A thesis submitted for the degree of a Philosophiae Doctor

Structure-Dynamics-Function Relation of Metalloproteins: Coupling Magnetic Resonance Spectroscopy and Simulation

---

Author:

Micha B. A. Kunze

Supervisors:

Dr D. Flemming Hansen

Prof Peter V. Coveney

Prof Christopher W. M. Kay

January 18, 2015

Declaration

I, Micha Ben Achim Kunze confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Micha Ben Achim Kunze

London, September 2014

Abstract

Dynamics of proteins are increasingly recognised as key features as they can contribute to the function of the protein. Structural dynamics manifest as protein folding, protein domain movement and small allosteric responses. Hence, investigating and understanding atomistic motions in proteins to elucidate their implication in protein function mark a crucial paradigm shift from a structure-function relation to a structure-dynamics-function relation.

The human histone deacetylase 8 is a key hydrolase in gene regulation and has been identified as a drug target for the treatment of several cancers. I used molecular dynamics simulations to propose a mechanism by which dynamic loop interactions can influence the activity of the human histone deacetylase 8. Subsequently I substantiated this hypothesis by using experimental techniques such as biochemical assays and single point mutations. Furthermore, I studied the structure and dynamics of the histone deacetylase 8 using nuclear magnetic resonance techniques.

The proposed mechanism of loop interaction yields a mechanistic rationale for phenomena that could not be explained on a molecular level before.

Contents

List of Figures

Preface

> "How would we express in terms of the statistical theory the marvellous faculty of a living organism, by which it delays the decay into thermodynamical equilibrium (death)? ... the device by which an organism maintains itself stationary at a fairly high level of orderliness ... really consists in continually sucking orderliness from its environment."
>
> Erwin Schrödinger, 1944

I embarked my Ph. D. studies with a B. Sc. in physics, which meant that most biological concepts were completely new to me. However, looking at biological processes from a physical perspective always posed fascinating and at the same time complex problems for me. This is exemplified in the above quote from Erwin Schrödinger in the preface of his book 'What is Life?'. Understanding biology or life using physical predictive models gives a plethora of interesting questions and riddles.

The quote of Erwin Schrödinger is as much motivation as it is a memento: It speaks about the 'marvellous faculty of living organisms', but also reminds me as a scientist that there is a discrepancy of studied 'equilibrium' model systems and the inherent 'non-equilibrium' features of living organisms. Hence, studying biological systems is never trivial and always reveals just another glimps of nature itself.

It was my great pleasure to work with many different great scientists during my time as a Ph. D. student and to be involved in a variety of projects.[1–5] My main research project was concerned with the investigation of functional dynamics of the human histone deacetylase 8 (HDAC8), which is the main focus of this thesis.[1]

1. Introduction

My interest in molecular modelling techniques and magnetic resonance has allowed to me to work on several projects in life[1–4] *and material science.*[5] *However, the main focus of my thesis is the elucidation of functional implications of protein dynamics on an atomistic scale.*

This chapter first gives a motivation to study functional protein dynamics and introduces the applied techniques. Subsequently the biological relevance and context of studying the human histone deacetylase 8 (HDAC8) are addressed.

1.1. Protein Dynamics

It has been established that protein structures are capable of sampling many different conformations, which can have an impact on the function of the protein.[6–14] It should however be noted that there is a debate in the field about how dynamics itself are influence protein activity, e.g. the catalytic step of an enzyme.[15] That is, whether dynamics, i.e. movements, influence the activity or dynamics generate different structures that give rise to a different activity.

Nevertheless, the dynamic nature of protein structures is key, as can be easily seen by the phenomenon of protein folding.[16] Such protein dynamics are conceptually often described using an energy landscape, a hyper-surface that describes the thermodynamic free energy as a function of the degrees of freedom.[16] A prominent example is the 'funnel shaped' energy landscape of protein folding. Similarly the sampling around the solution structure of a folded pro-

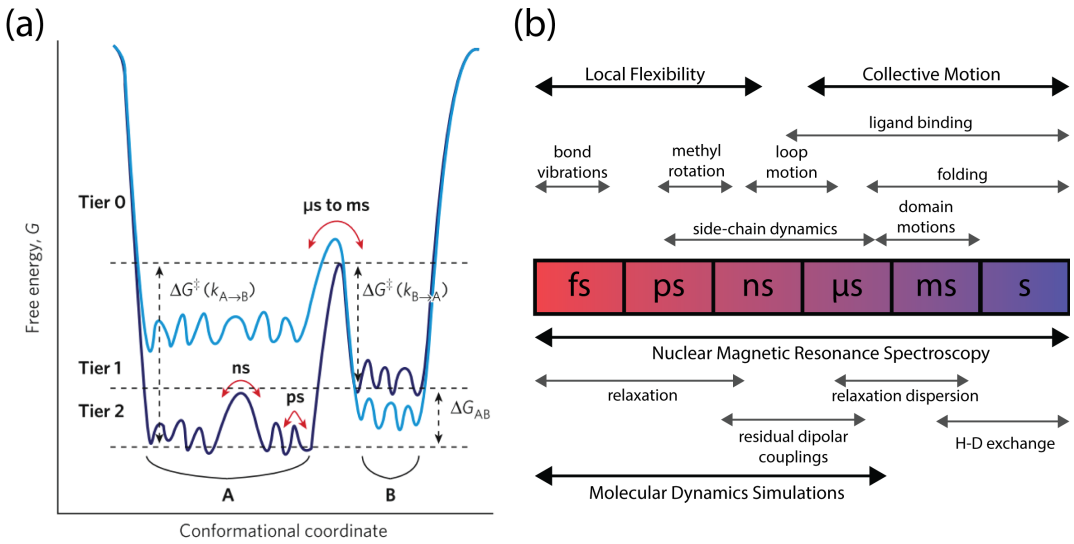


Figure 1.1.: (a) Slice of a hyper-dimensional energy landscape where minima in the energy function define a state and the height of the barriers define the time scales of the motion between the states. Motions between the states can be separated into Tier 0, 1 and 2 motions.[6] Changing the system (e.g. introducing a point mutation in a protein (Section 2.1)) alters the energy landscape - blue versus cyan. Figure adapted from Ref. 6. (b) Accessible timescales of NMR spectroscopy and MD simulations. Timescales are illustrated with a colour gradient going from red (femtoseconds) to blue (seconds). Dynamic molecular processes are highlighted within their corresponding timescale above the timescale gradient. Below the gradient the accessible timescale of NMR and MD are highlighted, including some specific experiment types.

tein can be drawn up in such an energy landscape and divided into different conformational states.[6]

Transitions between the states, i.e. closely related structures in an energy well, take place on different timescales implying different energy barriers. The transitions can be categorised by their characteristics (Figure 1.1 (a)). Tier 0 transitions are concerted motions of many amino acids separating kinetically distinct states of the protein (microseconds − milliseconds), i.e. states that are separated by several k_BT, where k_B is the Boltzmann constant and T is the temperature (States A and B in Figure 1.1 (a)). Tier 1 are e.g. loop motions

occurring within a Tier 0 state (nanoseconds) and Tier 2 are side-chain motions (picoseconds) or bond librations. Which of these motions and associated states are present and important for the function of a specific enzyme is generally not known. Any conformational transition of the different tiers could be important in a given case.[6,11,17,18]

It immediately becomes clear that in order to investigate any functionally crucial dynamics (Figure 1.1(b)), a toolbox has to be used that is capable of resolving the relevant time and space scales. Since loop flexibility seems to be a major component of the HDAC8 dynamics, as described later in Section 1.4.3, a combination of molecular dynamics (MD) simulations and nuclear magnetic resonance (NMR) along with other biochemical techniques to study important dynamic features of the enzyme seems extremely promising.

On the one hand MD allows observation of how the protein moves on the atomistic level on the fs $-$ μs timescale. On the other hand NMR is capable of measuring these motions in an experimental ensemble measurement. Hence, in combination these techniques allow a complementary observation of protein dynamics with an atomistic level of detail.

1.2. Introduction to MD

To study biomolecular motions theoretically MD simulations are commonly the method of choice and MD simulations are now widely used in biological research. Applications range from investigations of small molecule motions and free energy calculations over protein dynamics on timescales from nanoseconds[2,19] to, in favourable cases, milliseconds,[20,21] to the simulation of protein assemblies and whole viruses.[22]

As detailed below the underlying idea of MD is to use statistical mechanics to propagate the structural evolution of a given system in a thermodynamic ensemble.

1.2.1. Basic Concepts of MD

There are many different approaches for how MD simulations can be applied, all of which imply different approximations, i.e. the level of detail of the simulations can be tuned to fit the studied properties. In general, the bigger the system and the larger the amplitude of motions the larger the timescale of interest. Hence, for bigger systems more coarse-grained approaches are suitable, which yield enough detail for the studied motion, but omit enough of unnecessary details to speed up the computation.

Examples of this are fully atomistic simulations, taking each atom into account, or a more coarse-grained approach describing multiple atoms as one particle as the MARTINI force-field does.[23]

In my studies I used MD to simulate the structural motion of biomolecules on an atomistic scale and will hence stay in the picture of all-atom MD. However, classical all-atom MD can be viewed as a somewhat coarse-grained approach as well. Here, the atoms are assumed to be simple beads and forces are calculated from additive potential functions. Covalent bonds are treated as simple tethers, essentially 'coarse-graining' the whole electrons:nucleus interaction and anisotropies stemming from e.g. d-orbitals of transition metal ions.

These approximation of course come at the cost of several limitations, e.g. no bond breaking/forming can be simulated in classical MD and hydrogen exchange is neglected as well. While there are methods to incorporate such processes into an MD context, such as MD-REACT,[24] in the presented research where one is interested in atomistic motions, the influence of these processes can generally be neglected. Furthermore, this is the only way to achieve meaningful sampling and still keep atomistic resolution of the motions in the multi μs regime, which more detailed simulations, such as quantum chemistry and other *ab initio* techniques,[25] cannot achieve.

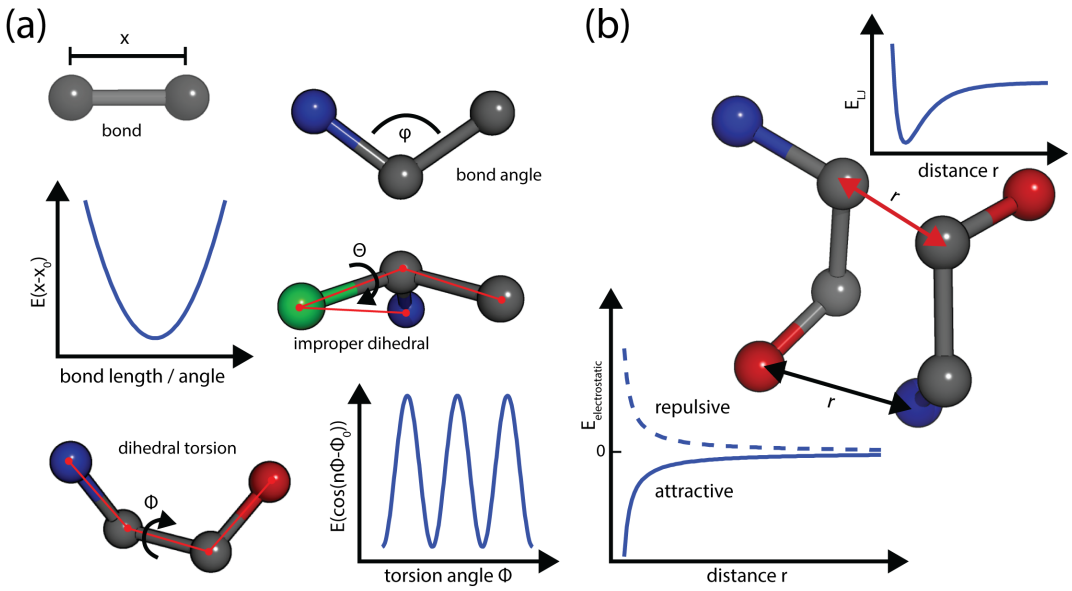


Figure 1.2.: Scheme of classical MD. (a) Modelling of covalently bonded atoms. Bonds are approximated by a harmonic potential, similar to the bond angles and improper dihedrals. Torsions are modelled using trigonometric potentials. (b) Illustration of non-bonded terms, the Lennard-Jones interactions (red arrow) and electrostatic interactions (black arrow).

Additive Potentials

In the context of all-atom MD, atoms are described as simple isotropic beads, covalent bonds are modelled as springs and dihedral and improper dihedral angles are modelled with a potential that describes their structural and dynamical behaviour. The consequence is that most details of the structure are coarse-grained and embedded into the additive potential model (Figure 1.2). For example, the covalent bond energy E_{bond} is approximated by describing the bond as a spring. Following Hooke's law ($F = -k\Delta x$, where F is the force acting on each atom when the spring is elongated by Δx) one has to find the right spring constant k to get the correct description for specific bond vibrations. This spring constant can be derived from experiments (e.g. infrared spectroscopy experiments) or from quantum chemistry calculations. While Hooke's law describes the force, the potential E_{bond} for such a bond is a simple

quadratic (harmonic) function. This can easily be seen by realising that the total differential of the potential is the force. In the simple one-dimensional case:

$$F = \frac{\mathrm{d}}{\mathrm{d}\Delta x} E_{bond} = \frac{\mathrm{d}}{\mathrm{d}\Delta x} \left(-\frac{1}{2} k \cdot \Delta x^2 \right) = -k\Delta x \tag{1.1}$$

This description originally stems from an harmonic approximation of the Morse potential[26] that describes the energy of a covalent bond of a diatomic molecule. The harmonic approximation (Formula 1.1) is only valid for small Δx, i.e. where the resulting bond energy is much smaller than the dissociation energy (bond-breaking energy), i.e. ideally $E_{bond} \ll E_{dissociate}$.

Other bonded parameters (E_{angle}, $E_{dihedral}$ and $E_{improper}$) can be described with potentials as well. For example, the energy of improper dihedrals is essential to maintain planarity in ring structures. Non-bonded interactions are simple additive potentials of van der Waals (vdW) interactions, which are approximated by a Lennard-Jones (LJ) potential (or in some cases a computationally more expensive Buckingham potential),[27] and the Coulomb energy.

The Lennard-Jones potential,[28] which is sometimes termed '12-6 potential', describes the interaction energy between two neutral atoms. Due to its simple form and computational efficiency this potential is widely used:

$$E_{LJ} = \varepsilon \left[\left(\frac{r_{min}}{r} \right)^{12} - 2 \cdot \left(\frac{r_{min}}{r} \right)^{6} \right] \tag{1.2}$$

Here, r is the distance between the atoms and r_{min} is the distance where the potential has its minimum, i.e. $E_{LJ}(r_{min}) = -\varepsilon$. The r^{-12} term is the repulsive term that stems from overlap of the electron clouds (Pauli repulsion) and the r^{-6} term is the vdW attraction force at longer range, i.e. the dispersive force. Of course r_{min} and ε can be determined from experiments or quantum chemistry calculations for the desired type of atom combinations.

The set of parameters describing such potentials including the non-bonded

electrostatic and vdW (or LJ) interactions is called a force-field, in case of the presented work a simple additive force-field is used:

$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral} + E_{improper} \quad (1.3)$$

$$E_{non-bonded} = E_{coulomb} + E_{LJ} \quad (1.4)$$

The parameters describe the energy landscape that each atom feels (E_{total}) and the total differential yields the force for time propagation by integrating of Newton's second law of mechanics. Many force fields have been developed and tested over many years[29] and it has been shown that the description of modern force-fields used is extremely powerful and in favourable cases can even be used to accurately describe the folding of small proteins *in silico*.[21,30,31]

Velocity Verlet Integration

Starting from an initial configuration, a set of coordinates $q(t = 0)$ and momenta $p(t = 0)$ (velocities $v(t = 0)$ are assigned using a Maxwell distribution), one can then solve Newton's classical equations of motion in the energy landscape and propagate the movement of the molecules. Solving Newton's equations for such a system, which can comprise millions of atoms can be handled by a range of algorithms, e.g. a Velocity Verlet algorithm.[32]

The Velocity Verlet consists essentially of two steps. First the position is propagated for the time increment Δt to time $t + \Delta t$ using the velocity v and acceleration a at the current position, i.e. at time t. In the second step the velocity is propagated based on the current velocity and acceleration and the acceleration at position $x(t+\Delta t)$, which has been calculated in step one. Thus, the Velocity Verlet algorithm can be summarised as:

$$\text{Step 1}: \quad x(t + \Delta t) = x(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2$$

$$\text{Step 2}: \quad v(t + \Delta t) = v(t) + \frac{a(t) + a(t + \Delta t)}{2}\Delta t$$

Hence the system can be propagated in the phase space by iteratively applying this algorithm. It is noteworthy that a sensible choice of the integration time step Δt is imperative for a stable propagation since the discretisation error scales with the size of Δt. The choice of the integration time step is system dependent and is a very important stability criterion. Generally, this should be much smaller than the timescale of the fastest motion in the system. For proteins the fastest motions are the heavy atom - proton bond vibrations, which are on the order of $10^{-14}\,\mathrm{s} = 10\,\mathrm{fs}$. Hence, a Δt of 1 fs is appropriate for a simulation involving these bond stretches. However, these bonds can be constraint, i.e. kept fixed, which a) eliminates this motion as the fastest motion and b) effectively removes a degree of freedom. I have made use of such constraints to be able to use an integration time step of 2 fs.

While the described principles build the basis for MD simulations, depending on the simulated thermodynamic ensemble and system, things such as pressure or temperature coupling or periodic boundary conditions need to be taken into account as well. However, for the sake of brevity they are not discussed in detail here.[33–35]

1.2.2. State-of-the-Art MD

The field of MD simulations has gained a lot of attention due to the 2013 Nobel Prize in Chemistry, which has been awarded to Martin Karplus, Michael Levitt and Arieh Warshel 'for the development of multiscale models for complex chemical systems' (www.nobelprize.org). The Laureates have been pioneers in the field of quantum mechanics / molecular mechanics (QM/MM) hybrid simulations. However, many other scientists, such as Michele Parrinello and co-workers should also be considered among the founders and pioneers of these methods.[25,36,37] They have also tremendously contributed to the development of many of the enhanced sampling methods described later.[38–43]

The recent development in the field yields sufficiently parallelised and opti-

mised MD simulations on modern computer clusters to make longer time scale motions, i.e. multi μs, accessible even on medium (10,000 to 100,000 atoms) sized systems. Here, not only central processing unit (CPU) power can be harvested, but more and more software is indeed harvesting the computing power of graphical processing units (GPUs).[44,45] The power and development of these GPUs is enormous, NVIDIA (Santa Clara, CA, USA) spent over a billion US $ on research and development of gaming GPU cards with over 4 billion US $ revenue in the 2014 fiscal year. Nearly 50% of this revenue is driven by the gaming industry, i.e. the revenue from the gaming GPU cards of NVIDIA (http://nvidianews.nvidia.com).

On top of this is the development of purpose-built supercomputers that are designed to perform MD simulations much faster than commodity hardware does. Prominent examples are the MDGRAPE supercomputer[46] or the Anton supercomputer,[47,48] which is approximately 1000 times faster per calculation step compared to commodity clusters.[47,48]

Sampling Issue

In parallel to the advances in software and hardware development the methodology of atomistic MD simulations has been developed as well. The standard approach of unbiased MD simulations is faced with what is often referred to as the sampling problem. This describes the ergodicity problem of such simulations, i.e. ergodicity is implied but not necessarily fulfilled as e.g. simulations get stuck in a local minimum or are just not extensive enough to sample all relevant parts of the phase space. The phase space is spanned by the coordinate q and momentum p of each configuration, where its population is described by the partition function:

$$Z = \sum_{n}^{N} exp(-\beta E_n(p,q)), \tag{1.5}$$

where $\beta = \frac{1}{k_B T}$ is the inverse temperature. E_n is the energy of the state depending on the potential energy U and the kinetic energy V:

$$E_n(p_n, q_n) = U(p_n) + V(q_n), \tag{1.6}$$

where n is the described degree of freedom and N is the total number of degrees of freedoms. For a simple problem, such as the Lenz-Ising model in a two dimensional lattice, the partition function can be solved analytically. In fact, in case of the Lenz-Ising model there are many papers with different solutions,[49] highlighting the complexity of this function even for simple models.

However, in systems containing tens of thousands of atoms the partition function cannot be solved analytically, which creates the necessity to use numerical tools, i.e. MD simulations. Hence, this is the root of the sampling problem: the numerical solutions of MD simulations are hard to test rigorously as the phase space and partition function are not known.

Long timescale simulations or replica simulations, i.e. the same system is replicated and simulations are carried out for each replicate using different starting configurations, are used to try to alleviate this problem. Unbiased replica simulation can also be guided by 'on the fly model building' using e.g. the Copernicus approach[50] to facilitate parallelisation on supercomputers with many cores.

Force Field Development

A key part of MD simulations is the force field that has to be accurate enough to give a good model for the phase space of a protein. Most force fields for atomistic MD rely on simple additive potential functions as the core modelling of atomistic interactions as described in Section 1.2.1.

There are several prominent force field branches for protein MD, which originated from different groups and MD software packages. Some of the most prominent are CHARMM from the Karplus group,[51] Amber from the Kollman

group[52] and GROMOS from the van Gunsteren and Berendsen groups.[53] However, there is continuous development of these force fields and water models such as TIP3P.[54]

The protein force fields have been developed over many years and are still being actively tweaked by the community. For example the Amber type ff99 has been modified several times; the ff99SB modification[55] improved the backbone dihedral terms, the ff99SB-ILDN[56] modification tweaked the side-chain torsion potentials of the Ile, Leu, Asp and Asn residues and the ff99SB-ILDN* modification by Hummer and Best[57] improved the helix-coil transitions via improvement of backbone potentials. In fact, these tweaks are not only applicable to the Amber type ff99 force field, and can in principle be used in other force fields, such as CHARMM22*, which performs well.[29]

However, not only protein force fields are actively developed. The water description has also been subject to development and tweaking. Water can be described with different numbers of particles, e.g. 3 as in the TIP3P model or 4 as in the TIP4P model.[54] These models have also been improved, such as the TIP4P-Ew for the use with Ewald summation techniques.[58]

For both, polarisable force fields have been developed, such as COS/D2 description for water[59] or the AMOEBA for proteins.[60]

However, using a water model that is modelled with more particles also increases the computational cost, which is also true for the use of polarisable force fields. Hence, the gained accuracy is a tradeoff for sampling.

Parallelisation Methods

Next to force fields the actual algorithms that make meaningful sampling of large systems possible are under constant development. There is a huge number of optimisations and hardware improvements taking place, including the use

of GPUs as described above. This also includes the improved use of CPU accelerations such as SSE2, SSE3 or techniques such as hyper threading.

However, inherently the bottle neck of MD simulations arise from the need to calculate long range forces where each particle is interacting with each other particle in the system. This of course creates problems for parallelisation on many CPUs as each process would in principle need to be constantly updated with the position of each particle in the simulated system, creating a lot of information that has to be passed between the CPUs at each step. Hence, several schemes have been developed to tackle this issue, which on first glance seems especially problematic for big systems with many particles.

An important aspect of efficient parallelisation are different cut-off schemes and the separation of short range and long range parts of interactions. For electrostatic forces the long range interaction is often calculated using PME techniques,[34] where the position of particles is only updated every certain number of steps (e.g. every 20 steps) between which the particles are considered stationary. This is a good approximation as for long ranges the actual movement of particles in a typical ambient MD simulations is rather small within e.g. 20 steps.

To make the calculations of short range interactions more efficient an approach called domain decomposition is used, which I also used when writing the ASD code described in Section 5.5. Here, the simulation cell is split into domains, which are individually simulated (for a certain number of steps). Each domain, or particles in each domain, only need information about what is happening within the short range cut-off of themselves. This can be done by using the domains and neighbouring domains or neighbour lists. While these need to be updated due to movement of particles between domains or neighbour lists, this can also be done only after a certain amount of steps similar to updating the particle information for long range interaction calculation.

Hence, even simulating large systems such as a whole virus[22] is possible using

CPU parallelisation. However, to be efficient there have to be a sufficient number of particles in each domain, i.e. a system can not be simulated faster by just using more and more cores. However, due to the (at least implied) ergodicity replica simulations of the same system can be run to increase the sampling of phase space using more CPUs than can be efficiently be used to simulate a single trajectory.

Enhanced Sampling Methods

There also have been methodological developments to enhance the sampling of phase space and thereby get closer to a true representation of a thermodynamic equilibrium system. A prominent example of these approaches is the metadynamics[13,37,42,61–63] method that uses additional variables, called collective variables (CVs), along which the sampling is enhanced by the addition of a history dependent potential function $V_{meta}(t, CV)$.

Thus, the sampling along the CV is monitored during the simulation and energy terms are added at the sampled positions until the system freely diffuses along the CV. The imprint of the added energy terms should ideally be a direct imprint of the sampled energy along this CV, allowing the measurement of the energy along the CV, i.e. $E(CV) = -V_{meta}(t, CV)$ for $t \to \infty$. The main drawback of this method is that the sampling depends on a good choice of CVs that accurately describe the processes involved and have no slower processes transverse to the chosen CVs. However, the quality of CV choice can at least be estimated from replicas where differences in the energy profiles can make hysteresis visible.

There are now many other approaches to enhance the sampling of MD simulations, such as accelerated MD in the AMBER package[44] or replica-exchange MD (REMD)[64] and a combination of well-tempered metadynamics[41,65] and the parallel-tempered metadynamics (PTMetaD - REMD with metadynam-

ics), the parallel-tempered metadynamics well-tempered ensemble (PTMetaD-WTE) simulations.[66]

The presented work focuses on long-time scale MD simulations that are intended for direct comparison to NMR data. In enhanced sampling methods, especially REMD type simulations, the time information is skewed or lost hindering a direct comparison of the experiment with experimental data, e.g. the direct calculation of order parameters via autocorrelation functions of bond vectors. As described in Section 2.1 I also applied enhanced sampling methods, i.e. metadynamics, as means of improving the sampling of interesting transitions and quantify the free energy profile of these.

1.3. Introduction to NMR

NMR is an extremely powerful spectroscopic tool to study the structure and spatial and temporal dynamics of proteins on an atomistic scale.[9,17,67–72] While NMR has been considered to be severely limited to the study of small molecules or proteins ($< 20\,\mathrm{kDa}$) the boundaries of this spectroscopic tool have been pushed to become one of the most versatile structural methods in biology. Today NMR can be used to study many different sizes of proteins, even 900 kDa complexes.[69]

As mentioned in Section 1.1 NMR is capable of resolving dynamics of protein structures on a range of timescales, i.e. it covers a broad range from picosecond to second,[6,73] mainly by measuring relaxation rates of different nuclei,[71,74] measuring residual dipolar couplings[75,76] or measuring relaxation dispersion.[7,8,70,77] These methods cover the range from ps to ms, for dynamics on the second timescale hydrogen-exchange experiments can be employed.[6,73] However, in the presented work side-chain motions have been measured with a focus on the ps to ns timescale for comparison with MD simulations. Hence, in this introduction only these type of experiments will be discussed.

In the following I will give a brief introduction to the concepts used in Section 2 without delving into the very basics of NMR spectroscopy. That is, I will assume that the theory of the Zeeman effect (chemical shift), scalar coupling of spins and the phenomenon of relaxation are familiar to the reader.

1.3.1. NMR of Biological Systems

NMR spectroscopy probes the nuclear spins of a molecule, which are influenced by their specific micro-environment and therefore resonate at specific frequencies, i.e. experience a chemical shift from the unperturbed resonance. These chemical shifts give a finger print of the molecule and can already yield information on the protein backbone conformation or secondary structure. The dispersion of a 1D ^{1}H spectrum can even hint at whether or not a protein is folded since an unfolded polypeptide has a much more narrow dispersion of peaks of e.g. the amide protons. Here, all spins sample an average environment, giving rise to an observed average chemical shift. As will be shown later, simple 1D ^{1}H NMR can also be used to differentiate small peptides such as the conversion from acetylated to de-acetylated peptide as shown in Section 2.3.[1]

However, in the context of structural biology one is more interested in the actual three dimensional structures of proteins. NMR can be used to solve such structures in solution, i.e. with the protein in aqueous solution even at ambient temperature.[72,78] This can be achieved by combination of experiments reporting on distances, bond angles, dihedral angles etc., where the combined information is finally used as constraints to calculate the structure or an ensemble of structures,[72] which are often called models.

Insensitive Nuclei Enhanced by Polarisation Transfer

An important aspect of biological NMR is the use of multi-dimensional experiments. Here, the dimensions are the different nuclei or coherences that are

targeted in the specific experiment, i.e. magnetisation is transferred between coherences of different spins/nuclei.

As an example of how magnetisation is transferred between nuclei the insensitive nuclei enhanced by polarisation transfer (INEPT) is briefly discussed. This scheme is central to most multi-dimensional NMR pulse sequences, not only for magnetisation transfer, but for the enhancement of magnetisation of insensitive nuclei as the name suggests. Equilibrium magnetisation is proportional to the energy difference between the spin states ΔE, which is given by the Zeeman splitting $\Delta E = h\gamma B_0$ in the high-field approximation, where h is Planck's constant. Hence, for a simple spin 1/2 nucleus the equilibrium magnetisation is proportional to the gyromagnetic ratio γ of the nucleus. The larger the magnetisation the larger the signal, making high polarisation desirable. By transferring magnetisation from a nucleus with a high gyromagnetic ratio to one with a small gyromagnetic ratio one can therefore enhance the polarisation and subsequently the signal.

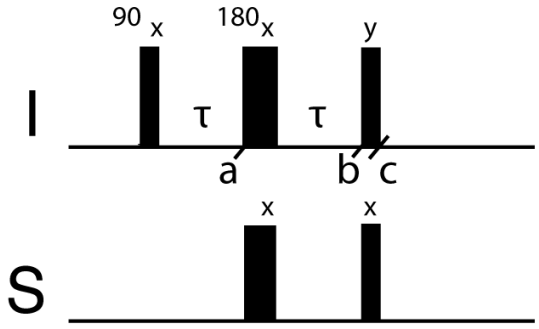


Figure 1.3.: Pulse scheme of an insensitive nuclei enhanced by polarisation transfer (INEPT). Narrow (wide) blocks depict 90° (180°) pulses along a specified axis (x, y) for spins I at the top line and S at the bottom line. Points a-c are guides for the magnetisation evolution discussion in the text.

In the case of a $^{1}\mathrm{H}-^{15}\mathrm{N}$ spin-pair, e.g. an isotopically labelled amide of a protein backbone, the ratio of the gyromagnetic ratios is:

$$\left|\frac{\gamma_{^1\mathrm{H}}}{\gamma_{^{15}\mathrm{N}}}\right| = \left|\frac{42.6\,\mathrm{MHz\,T^{-1}}}{-4.3\,\mathrm{MHz\,T^{-1}}}\right| = 9.9, \tag{1.7}$$

which means that a transfer from a proton spin to a nitrogen enhances the magnetisation of the nitrogen spin almost 10-fold, neglecting diminishing contributions such as relaxation processes during transfer. It is noteworthy that while the gyromagnetic ratio of $^{15}\mathrm{N}$ is negative, this has no effect on the transfer, only on the direction in which the magnetisation of this nucleus precesses.

The magnetisation transfer during an INEPT can be easily understood by the use of the product operator description.[79] In the general case of two spins I and S, which both have a positive gyromagnetic ratio $\gamma_I > \gamma_S > 0$, that are weakly coupled by the scalar coupling J_{IS} Figure 1.3 shows the INEPT pulse scheme. Assuming that only I_z equilibrium magnetisation is present in the beginning of the sequence and neglecting free precession as this is refocused with the $\tau - 180° - \tau$ sequence, the first $90°_x$ pulse on the I spin produces $-I_y$ magnetisation. The product operator description gives:

$$I_z \xrightarrow{\frac{\pi}{2}I_x} I_z \cos\frac{\pi}{2} - I_y \sin\frac{\pi}{2} = -I_y$$

$-I_y$ evolves under the scalar coupling Hamiltonian $H = \pi J_{IS} 2 I_z S_z$ over a time τ, neglecting free precession as described earlier. Hence at point a) the following magnetisation is present:

$$\xrightarrow{\pi J_{IS}\tau 2 I_z S_z} -I_y \cos(\pi J_{IS}\tau) + 2 I_x S_z \sin(\pi J_{IS}\tau)$$

The following I and S 180°_x pulses invert the magnetisation of both spins that are not along x:

$$\xrightarrow{\pi I_x, \pi S_x} I_y \cos(\pi J_{IS}\tau) - 2I_x S_z \sin(\pi J_{IS}\tau)$$

After the pulses, scalar coupling is allowed to evolve:

$$\begin{aligned} &\xrightarrow{\pi J_{IS}\tau 2I_z S_z} I_y \cos^2(\pi J_{IS}\tau) - 2I_x S_z \cos(\pi J_{IS}\tau)\sin(\pi J_{IS}\tau) \\ &\quad -2I_x S_z \sin(\pi J_{IS}\tau)\cos(\pi J_{IS}\tau) - I_y \sin^2(\pi J_{IS}\tau) \\ &= I_y \underbrace{\left(\cos^2(\pi J_{IS}\tau) - \sin^2(\pi J_{IS}\tau)\right)}_{=\cos(2\pi J_{IS}\tau)} - 2I_x S_z \underbrace{2\left(\cos(\pi J_{IS}\tau)\sin(\pi J_{IS}\tau)\right)}_{=\sin(2\pi J_{IS}\tau)} \end{aligned}$$

yielding $I_y \cos(2\pi J_{IS}\tau) - 2I_x S_z \sin(2\pi J_{IS}\tau)$ magnetisation at point b). The following $90^\circ I_y$ pulse does not effect the $I_y \cos$ component, as I_y and I_y commute, i.e. $[I_y, I_y] = 0$. However, it transfers the transverse magnetisation I_x of the $2I_x S_z \sin$ component into longitudinal magnetisation. The $90^\circ S_x$ pulse produces transverse S_y magnetisation, yielding $I_y \cos(2\pi J_{IS}\tau) - 2I_z S_y \sin(2\pi J_{IS}\tau)$ magnetisation at point c). By choosing τ as to maximise the $S_y \sin$ term, as one is interested in the produced transverse S_y magnetisation, the complete magnetisation of I can be transferred in this ideal case, i.e. $\tau = \frac{1}{4J_{IS}}$. Hence, the overall effect on the magnetisation of the INEPT sequence can be summarised as:

$$I_z \xrightarrow{INEPT} -2I_z S_y \tag{1.8}$$

Formula 1.8 shows that magnetisation can easily be transferred between weakly coupled spins, which is the case for nuclei found in (isotopically labelled) proteins, i.e. ^{1}H, ^{2}H/D, ^{13}C and ^{15}N.

Multi-dimensional NMR

In a simple two-dimensional $^{1}H - ^{15}N$ experiment of for example the amide hydrogen and amide nitrogen of an isotopically labelled protein backbone, magnetisation is transferred from the proton to the nitrogen. Subsequently the magnetisation is allowed to evolve under free precession, i.e. the magnetisation gets modulated with the chemical shift of the nitrogen nucleus in an indirect dimension. Magnetisation is then transferred back to the proton where the free induction decay (FID) is detected, allowing the direct measurement of the proton chemical shift. By recording several increments in the indirect dimension, i.e. increments in the evolution time of the free precession of the nitrogen magnetisation, a data set is recorded that entails information about each proton spin (direct detection of the FID) and the scalar coupled nitrogen spin (indirect dimension), essentially creating a two-dimensional spectrum where the correlation map of proton and nitrogen spin can be observed by the cross peaks at the specific frequencies.

In a similar fashion three- and higher dimensional NMR spectra of proteins can be recorded, e.g. with an HNCO experiment. HNCO describes the magnetisation pathway (from the proton of the amide 'H' to the nitrogen of the amide 'N' to the carbonyl carbon 'C'). This and other similar spectra yield information about the connectivity of amino acids. Other experiments such as Nuclear Overhauser Effect (NOE) measurements yield distant constraints for the final structure calculations. Obviously, NMR is very powerful tool to investigate the structure of proteins. However, there are still several factors which make its application not always straight forward.

Limitations of Biological NMR

Limitations include the need for purified isotopically labelled protein, which often poses problems for proteins of eukaryotic origin, especially when the

protein structure relies on disulphide bonds, needs specific chaperons or is heterogeneous in nature such as glycoproteins.[80] However, these problems can in some cases be alleviated, e.g. by optimisation of expression conditions to facilitate disulphide bond formation[72] or reducing the heterogeneity.[80]

Another aspect is the sample concentration, which should in most cases ideally be $> 200\,\mu$M in a volume of at least $250\,\mu$L for a good signal-to-noise ratio of spectra recorded in a reasonable time, although simple 2D spectra can be obtained for samples with a concentration as low as $5\,\mu$M. In addition, the studied biomolecule should be stable for a long time, i.e. the protein must not aggregate in a few hours after the purification and should preferably be stable for several days or weeks at ambient temperature. This is necessary as many experiments, especially higher dimensional and relaxation experiments, can take several days of acquisition time.

Regarding the size of the studied system there are two major aspects that cause difficulties for the application of NMR. The first is spectral overlap in multi-dimensional NMR spectra. The bigger the protein the more peaks there are, increasing the chance that these overlap. This overlap results in difficulties for the assignment and analysis of the peaks up to a point where the interpretation of the data becomes impossible. The other major aspect is the phenomenon of relaxation. Transverse relaxation increases with molecular weight, leading to signal loss and line broadening that causes difficulties for proteins larger than 40 kDa at current magnetic fields using standard experiments.[81]

To tackle the difficulties arising from spectral overlap and relaxation many approaches have been developed. An obvious approach is to work at higher frequencies, which is limited by hardware, and typical spectrometers nowadays operate between 400 and 1000 MHz. Another way to reduce spectral overlap is to use higher dimension NMR or clever labelling schemes such as amino acid specific isotope labelling, i.e. only a certain amino acid or a set of amino acids are labelled with isotopes such as ^{13}C or ^{15}N. A prominent example

is the isotopic labelling of methyls of the residues isoleucine, leucine and valine.[82–84] Amino acid specific labelling is typically achieved by addition of a specific precursor,[85,86] in Section 2.4 I used such an approach to specifically label the $C_{\delta 1}$ position of isoleucine side-chains. To study larger proteins where the transverse relaxation contributes significantly to signal loss, pulse schemes have been developed where the transfer of magnetisation is relaxation-optimised,[82] i.e transverse relaxation-optimised spectroscopy (TROSY).

I used isotope labelling of the methyl group of the $C_{\delta 1}$ of isoleucines to study the dynamics of a 42 kDa metalloenzyme employing TROSY type pulse schemes. Hence, I will in the following focus on methyl TROSY and the basic concepts of how to measure dynamics of proteins with NMR.

1.3.2. Methyl TROSY

TROSY sequences rely on selecting specific transitions or coherences of a spin system that have favourable relaxation properties, i.e. typically a slow relaxation. This is a) to keep as much magnetisation as possible as it relaxes during the acquisition sequence and b) to measure slow relaxing coherences to obtain sharp lines as the width of a peak is generally governed by the effective relaxation rate R, the decay of the measured FID. The line width $\Delta\nu_{1/2}$ of a peak is proportional to R. The differences in relaxation rates of the different coherences arise from the combinations of relaxation contributions. Those contributions can add up or cancel each other. In the $^{1}H-^{15}N$ example from earlier the TROSY effect is present due to favourable combination of the chemical shift anisotropy (CSA) and dipolar coupling interaction, which give rise to a transition that relaxes much slower than the others.

However, in my work I used methyl TROSY experiments. For a $^{13}CH_3$ isotopomer the transverse relaxation optimisation comes naturally when using proton carbon heteronuclear multiple quantum correlation (HMQC) experiments (Figure 1.4). This is different from heteronuclear single quantum corre-

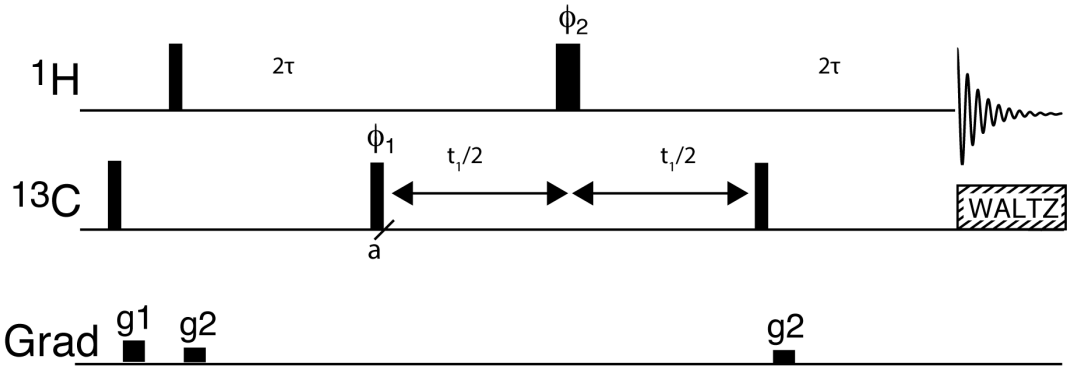


Figure 1.4.: HMQC pulse scheme for methyl TROSY. Narrow (wide) blocks represent $\frac{\pi}{2}$ (π) pulses and are applied along the x-axis unless otherwise noted. Details for the sequence can be found in Ref. 82.

lation (HSQC) experiments as during the carbon magnetisation evolution proton magnetisation is allowed to evolve, thereby producing multiple quantum coherences. Figure 1.4 shows such a HMQC pulse scheme for the acquisition of $^{1}\mathrm{H} - ^{13}\mathrm{C}$ 2D correlation maps.

In a methyl group the relaxation is mainly governed by dipole interactions. However, similar to the situation of the amide $^{1}\mathrm{H} - ^{15}\mathrm{N}$ correlation where CSA and dipolar field cancel each other, there are certain dipolar interactions that cancel in methyl group and therefore allow for a TROSY type coherence, i.e. cancellation of cross- and auto-correlated relaxation terms.[82]

Essentially, at point a) in the sequence (Figure 1.4) during t_1 evolution in the un-decoupled case a 1:2:1 triplet is present in the indirect dimension. It can be shown, that the middle line is composed of transverse multiple quantum coherence (H_+C_+) and that this transition does not relax due to intra-methyl contributions, at least in the $J(0)$ limit.[82,87,88] It is however still effected by spin-flips caused by external protons, hence spectra should be recorded with a highly deuterated background to ensure a good signal-to-noise improvement (up to 3-fold[88]) over typical HSQC experiments. In an HSQC experiment, slow

and fast relaxing transitions are mixed, which especially in big proteins with large transverse relaxation rates R_2 leads to significant signal loss.

In general, for the application to methyl groups, an HMQC experiments is preferable. However, this will always depend on the specific labelling scheme and level of deuteration. In Section 2.4.2 this is discussed in more detail for the application on $^{13}CH_2D$ methyl isotopomers.

1.3.3. Studying Protein Dynamics using NMR

As described in Section 1.1 the dynamics of a protein structure can be important for the protein function. NMR is not only capable of reporting on the structure of a protein, but also its dynamics. The reorientation of molecular interaction vectors, e.g. dipole interaction vector, cause the local magnetic field to fluctuate and thereby contribute to the relaxation of the spin coherence.[89,90] By measuring the relaxation rates one can therefore probe the dynamics of a given vector, e.g. a bond vector.

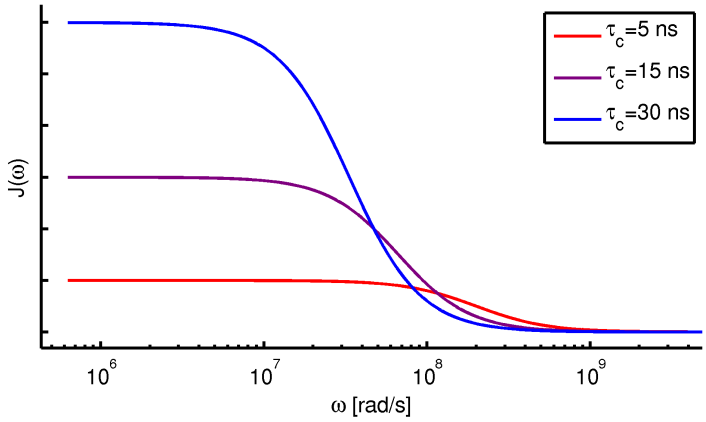


Figure 1.5.: Spectral density $J(\omega)$ plot for rotational diffusion with different correlation times: 5, 15 and 30 ns, coloured in red, purple and blue respectively.

This is generally done by mapping the spectral density function $J(\omega)$.[89,90] Phenomenologically, this function describes the probability of finding motions at a given angular frequency ω and is strictly given by the Fourier transform of

the correlation function of the motions. For a simple rotational diffusion the spectral density function takes the form:

$$J(\omega) = \frac{\tau_c}{1 + (\omega\tau_c)^2}, \tag{1.9}$$

where τ_c is the tumbling correlation time. Figure 1.5 shows what the spectral density for rotational diffusion as described in Formula 1.9 looks like. The red plot shows the spectral density for fast tumbling with a correlation time of 5 ns (approximate τ_c of a 10 kDa protein at 25 °C) and the blue plot for slow tumbling 30 ns (approximate τ_c of a 50 kDa protein at 25 °C). The difference becomes immediately clear for small ω values, i.e. for $J(0)$, which directly reflects on the timescale of the tumbling. Moreover, the inflection point is at higher ω values for faster tumbling. As can be seen in Section 2.4.2, knowledge about the relation of the relaxation rates and the spectral density lets one map different points of the spectral density, e.g. $\omega = 0, \omega_H, \omega_N$ for certain amide relaxation measurements or ω_D for deuterium relaxation measurements (Figure 1.6).

In the case of bond vectors in proteins the motion can be described by a spectral density function that has two (or more) components within the spectral density function. Lipari and Szabo have established this for the general case in their description of a model free approach to analyse relaxation parameters from NMR data.[89,90] Model free means that the used spectral density function does not imply a physical model of the motion, e.g. motion in a cone.[91] Taking the simplest two parameter approach the spectral density is described by two contributions, the first is the overall tumbling of the protein and the second the internal motion, which is what one is usually interested in:

$$J(\omega) = \frac{2}{5}\left(\frac{S^2\tau_c}{1 + (\omega\tau_c)^2} + \frac{(1 - S^2)\tau}{1 + (\omega\tau)^2}\right), \tag{1.10}$$

where $\tau^{-1} = \tau_c^{-1} + \tau_f^{-1}$ and is the effective internal motion correlation time and

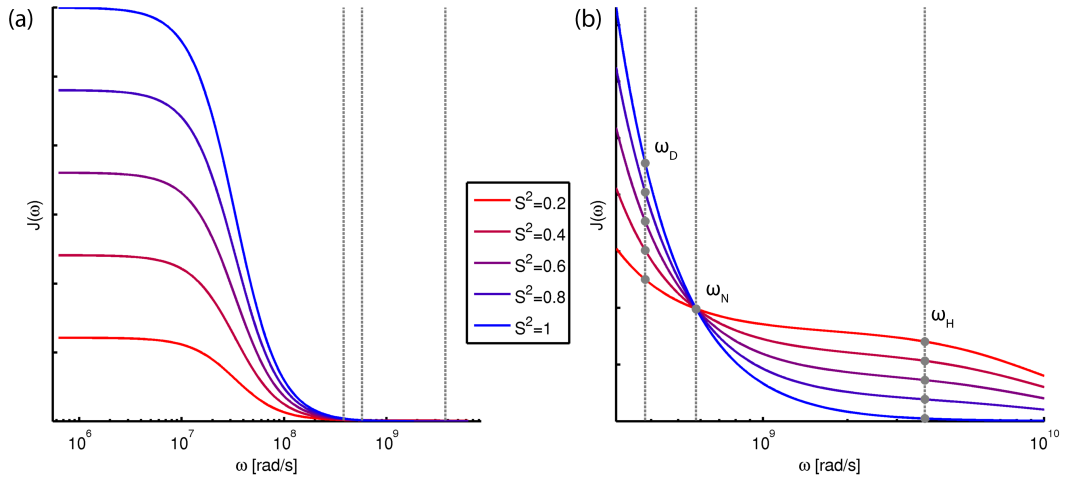


Figure 1.6.: Spectral density $J(\omega)$ plot for the simple LS2 model. Parameters for the plots are: $\tau_c = 30\,\text{ns}$ and $\tau_f = 100\,\text{ps}$ for different S^2 values from 0.2 to 1 coloured from red towards blue. Panel (a) shows the full range, where $J(\omega)$ for low ω values ($\omega < 10^6\,\text{rad}\,\text{s}^{-1}$) is converging to the $J(0)$ value. Grey lines indicate Larmor frequencies of measurable nuclei, where ω_H is the proton Larmor frequency, ω_D is the deuterium Larmor frequency and ω_N is the nitrogen (^{15}N) Larmor frequency. (b) Zoom in of (a) for the region close to the different Larmor frequencies. Grey dots illustrate how a spectral density can be mapped at different frequencies.

τ_f is the correlation time of the fast internal motion. The simple two parameter model is often called LS2 model, the Lipari-Szabo two parameter model. The contributions from the protein tumbling and the internal motion are separated by a generalised order parameter S^2, where the protein tumbling τ_c is much slower than the internal motion τ_f so that they are separable. S^2 then reports on how restricted the internal motion is, i.e. $0 < S^2 < 1$, where a small S^2 describes a motion that is less restricted (right term in Formula 1.10) and a high S^2 value more restricted motion (left term in Formula 1.10). However, S^2 is not a measure of the timescale of the motion and does not discriminate between different motions such as discrete jumps or continuous rotation.[89,90]

The LS2 model is described in more detail in Section 2.4.2, where it is used to fit methyl order parameters.

Backbone and Side-Chain Dynamics

As mentioned earlier, measuring relaxation rates and mapping the spectral density function at different points lets one obtain a generalised order parameter S^2 that reports on the restriction of motion on the measured timescale, i.e. a timescale shorter than the overall tumbling of the protein. It is noteworthy that such an order parameters can be obtained for different bond vectors. A prominent example is the $^{1}\mathrm{H} - ^{15}\mathrm{N}$ amide bond-vector to get a measure of the backbone dynamics and flexibility.

While backbone dynamics give a measure of the overall dynamics of the protein structure, the dynamics of side-chains can be of interest, especially the dynamics of charged residues.[92] It is noteworthy that backbone and side-chain dynamics are not generally correlated, hence if the side-chain is functionally important the dynamics of this should be studied. There has been a lot of interest in the application of methyl NMR techniques such as TROSY NMR to study methyl groups of aliphatic residues. Methyls yield good signals even in large proteins as they give a very good entry point for magnetisation. The protons are magnetically equivalent due to the fast rotation and they do not exchange with solvent protons, which is a significant advantage over amide protons. Moreover, aliphatic residues are quite abundant, which makes them well dispersed reporters on most proteins and can be effectively used to reduce spectral overlap.

1.4. Histone Modifications

Post-translational modifications of proteins are present in numerous signalling pathways in eukaryotic cells.[93,94] Especially for gene regulation, post-translational modification of amino acids of N-terminal histone tails was shown to be key for gene regulation and epigenetic landscaping.[93,94] Histones - H2A H2B H3 and H4 - form higher oligomeric structures which are able to pack DNA and

organise into the higher order structure of chromatin (Figure 1.7). Residues of N-terminal histone tails are subject to acetylation, mono- di- and tri- methylation, phosphorylation and ubiquitination.[93,94]

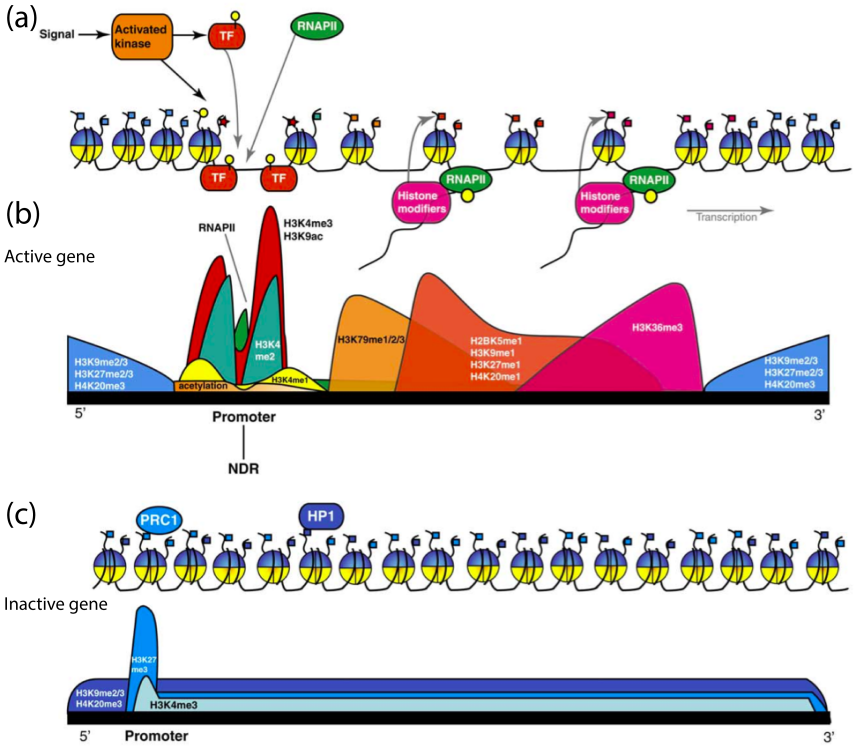


Figure 1.7.: Illustration of the different of modification pattern of actively transcribed genes (a-b) and inactive genes (c-d). (a) and (c) show a schematic view of the nucleosome and (b) and (d) show the post-translational modification pattern of a corresponding gene. This summarises that methylation is used as long-term marker and acetylation is a fast signal that leads to transcription. Figure reproduced from Ref. 93.

Despite direct influences on the properties of the histone tails, e.g. charge regulation of the tails via acetylation and deacetylation of lysines, which changes the interaction with negatively charged DNA, there are many proteins involved that sense the state of these key residues and interact with the gene expression apparatus, e.g. bromodomain-containing proteins.[95–98] It was suggested that the pattern of modification determines gene expression activity, as e.g. acety-

lation of lysines was associated with actively transcribed genes, and thus was termed ‘the histone code’.

However, it became clear that different modifications have very different half life values ranging from a few minutes in the case of acetylation (2 - 40 min)[99,100] to hours in the case of phosphorylation (30 min - 3 h)[99] up to days in the case of methylation (0.3 - 5 d),[101] which led to speculation that not only the static pattern of modification is important, but that modulations with a fast turnover rate act as signals. In this regard the concept ‘histone code’ is often expanded and referred to as histone ‘cross talk’.[93,94]

It is noteworthy that differences in life times seem to play a role in epigenetic landscaping as e.g. the methylation pattern is copied upon cell division. However, the dynamic aspects of histone modification and their implication on the theory of gene expression remain a matter of debate in the scientific community.[93,94]

The above demonstrates that gene regulation: a) is a dynamic and highly regulated process with many interaction partners, b) has many layers of complexity from local modifications over protein signalling to complete chromosome remodelling and c) is one of the most important topics in life science due to its implication on understanding life and treating diseases associated with gene regulation, i.e. cancer.

A crucial aspect of these concepts is that each modification needs an apparatus to maintain the correct pattern and dynamic turnover of modifications to maintain and regulate a healthy gene repression/expression. Thus, for each PTM there is a modifier for the residue and an antagonist that removes the modification. As can be seen from the above, histone tail modification is highly regulated, which in turn means that the apparatus that controls the modification must be highly regulated as well. This means that one needs to understand how histones and other proteins are modified and how modifications are removed to understand this crucial mechanism in gene regulation.

Acetylation of histone tail lysines was one of the first modifications that was found in conjunction with gene regulation and many studies have been conducted to investigate the role of this modification, its removal and the proteins that are associated with these modification and removal processes.[102–105]

1.4.1. Histone Deacetylases

Acetylation of N-terminal lysine side chains in histones is controlled by histone acetyl transferases (HATs), which acytelate lysines, and their antagonists: histone deacetylases (HDACs), which deacetylase acetyl-lysine.[106] Gcn5, which is a HAT, has early been found to be a transcriptional co-activator in yeast, which connected acetylation of such lysines to regions of active gene transcription.[107] Since then acetylation and deacetylation of N-terminal histone lysines where studied extensively.[108–111] It was recognised that the level of HDAC expression was altered in several cancers,[108] leading to the idea to target this post-translational modification as a novel cancer therapy.[95,105,112–114]

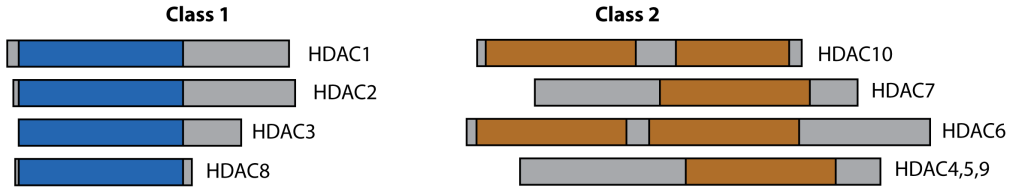


Figure 1.8.: Class 1 and 2 HDAC domain organisation. Highly conserved catalytic domains among each class are shown in coloured bars, blue (class 1) and orange (class 2), and less conserved auxiliary domains are shown in grey. Class 1 HDAC isoforms 1,2 and 3 share >43% sequence identity and >60% sequence similarity with HDAC8 (EMBOSS http://www.ebi.ac.uk/Tools/emboss/).

HDACs can be largely ordered into four major classes based on phylogenetic analysis and sequence similarity. Class I, II and IV are zinc-dependent HDACs with a conserved catalytic domain (Figure 1.8). HDAC8 is part of class I, which also comprises HDAC1, -2 and -3.[102]

Especially class I HDAC isoforms were found to be key entities in different areas of cancer cell biology, e.g. proliferation (HDAC1, 2, 3 and 8), apoptosis (HDAC1 and 2) and resistance to chemotherapy (HDAC1).[115] Moreover, it was found that expression levels of HDACs alter in certain cancers.[108] Thus, efforts have been made to identify suitable inhibitors for these enzymes, where two inhibitors have been approved by the Food and Drug Administration (FDA) of the United States of America, Vorinostat (SAHA - suberoylanilide hydroxamic acid) and Romidepsin, which are approved for the treatment of cutaneous T-cell lymphoma.[116]

1.4.2. Disease Implications of HDAC8

HDAC8 and class I HDACs in general have first been identified as cancer therapy targets.[112,117] HDAC8 in particular has been linked to cell differentiation and proliferation of cancer cells.[115] Recent studies show that it is of paramount importance for cancer treatment via HDACs to rather regulate individual isoforms instead of completely inhibiting one or all of them, due the very different functions and localisations of the HDAC isoforms.[95] This is exemplified by the correlation of HDAC8 activity with p53 expression levels.[95] p53 is one of the most studied tumour suppressors and is often called the 'guardian of the genome' as it regulates the cell cycle. A decreased level of HDAC8 leads to a decreased level of p53, however, effects are different for wild type and mutant p53. Again, this shows how important a detailed understanding of the gene-regulation machinery is for the development of anti-cancer treatments and therapeutics. In our research we focus on HDAC8, as it a) is a biomedically important target and b) mainly consists of the catalytic domain conserved in class I HDACs (Figure 1.8), making it a prime model system to draw general conclusions about HDACs of the same class.

Despite all efforts to characterise HDACs in detail the endogenous substrates of HDAC enzymes are still not unambiguously identified, which lead to HDACs

sometimes being called KDACs to signify their lysine (K) deacetylation (DAC) activity and not imply histones as their sole endogenous substrates.[96,118,119] HDAC8 can be found in the cytosol as well as in the nucleus and it seems to be capable of interacting with non-histone proteins. HDAC8 has been found to interact with the estrogen related receptor alpha[120] (ERR-α) and the structural maintenance of chromosome 3 protein (SMC3)[121] amongst others.[118,119,122]

In addition to cancer HDAC8 has recently been implicated in Cornelia de Lange Syndrome (CdLS)[121] via its deacetylase activity on SMC3. CdLS is a dominantly congenital inherited malformation disorder. Deardorff et al. found patients where HDAC8 mutations resulted in phenotypes that are similar to the classical CdLS without having the classical mutations of the cohesion loading protein NIPBL or components of the core cohesin complex.[121] Instead, mutations in HDAC8 were found that lead to a diminished deacetylation of SMC3 during anaphase causing gene expression patterns that are similar to CdLS related mutations of NIPBL.[121] Some of these mutations only down-regulate the enzymatic activity between 2 and 5-fold *in vitro* and still cause the same phenotype, showing that these small dis-regulations can have an enormous physiological effect.[121,123]

While some of the CdLS missense mutations can be readily explained, such as His180Arg as His180 is coordinating the catalytic Zn^{+2} ion, many of them are far away from the active site. Hence, these findings again sparked interest in functional mechanisms that these residues might be involved in.[121,123,124]

1.4.3. Structural Features of HDAC8

HDAC8, as all class I HDACs, is a zinc-dependent hydrolase and has been subject to extensive structural and biochemical studies, which provide a wealth of information on the structure of HDAC8 bound to different ligands.[125–129] The active site of HDAC8 consists of a divalent metal ion coordinated to a histidine (His180) and two aspartic acids (Asp178 and Asp267) (Figure 1.9).

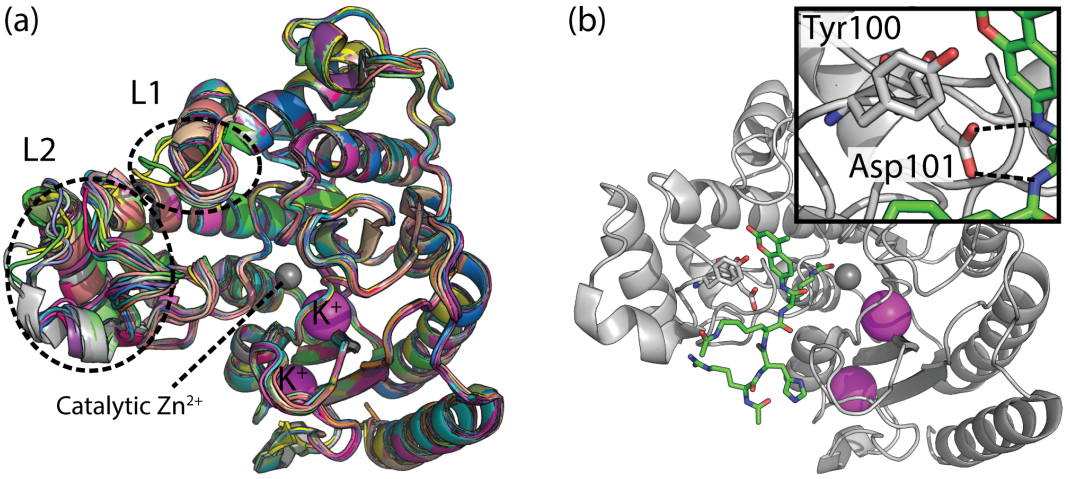


Figure 1.9.: Cartoon representation of HDAC8 crystal structures. (a) Overlay of HDAC8 crystal structures, loop L1 and L2 are highlighted. Position of zinc is shown as a grey sphere and position of potassium ions as magenta spheres. (b) Crystal structure of the HDAC8 Tyr306Phe mutant (grey) bound to substrate (green sticks). Inset is a zoom on the interaction of Tyr100 and Asp101 with the substrate. Figure adapted from Ref. 1.

Interestingly, it has been demonstrated that HDAC8 remains active with other divalent metal ions in the active site, such as the paramagnetic metal ions Co^{2+}, Fe^{2+} or Ni^{2+}, in the case of Co^{2+} activity is in fact increased 9-fold.[129] This metal site is on the bottom of a conserved entrance tunnel, which is surrounded by a network of loops. Many of these loops, in particular the L1 and L2 loop (Figure 1.9 (a) and Figure 1.10), are found in very different conformations in different crystal structures. It is noteworthy that, to the best of my knowledge, there has been no crystal structure reported for HDAC8 or any other class I HDAC without any ligand complexed. However, there is one structure in an asymmetric unit of three structures where no electron density could be found for an inhibitor,[127] yet no good quality crystals seem to form without addition of a ligand.[127] Crystal structures of HDAC8 exhibit a very conserved binding configuration formed by residues Tyr100 and Asp101,[125,126] which will be referred to as the 'binding rail'.

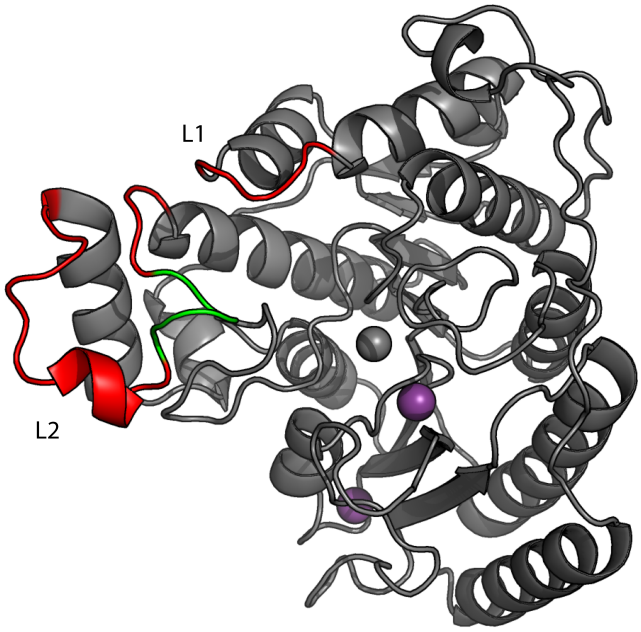


Figure 1.10.: Cartoon representation of HDAC8 crystal structure (PDB id 1T64). Loop 1 and 2 are highlighted in red and labelled. Both loops exhibit very little sequence similarity to other class I HDACs (1,2,3), expect the around the binding rail, which is located within the L2 loop and coloured green.

Specifically, Vannini et al.[126] showed that Asp101 forms hydrogen bonds with the backbone of a substrate thereby positioning and stabilising the substrate for the deacetylation reaction, as highlighted in the inset of Figure 1.9 (b). When a substrate-like ligand is bound, Tyr100 and Asp101 stabilise the binding via hydrogen bonds or stacking of aromatic groups. Moreover, mutating Asp101 to Ala renders HDAC8 inactive, which attests the importance of this residue and the binding rail in general for the enzymatic activity.[126] Other residues in the L2 loop are not conserved and the same is true for residues in the L1 loop, i.e. these regions exhibit the highest structure and sequence variations along the HDAC isoform and other orthologues.[125]

The L2 loop (Figure 1.10), which consists of residues 83 to 108, has two parts; residues 83-93 that are distal to the entrance tunnel and residues 93-108 that are proximal to the entrance tunnel and include the binding rail, Tyr100 and

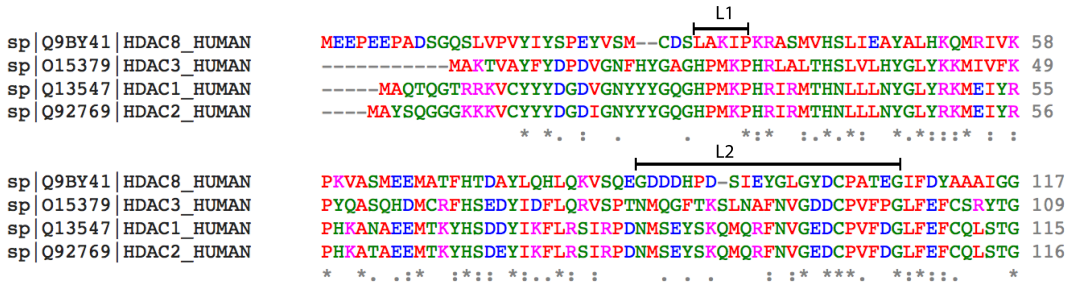


Figure 1.11.: Sequence alignment of the L1 and L2 region of class I HDACs (1,2,3 and 8), done using Clustal Omega (EMBL-EBI). Colours represent the residue type and the degree of conservation is indicated under the alignment (in descending order * : .). The position of the L1 and L2 loops of HDAC8 are highlighted and labelled. It can be seen that there is nearly no conservation in the loops and in fact loop insertions in the L1 in HDAC 1,2,3. The great exception is the region of the binding rail around residue Asp101.

Asp101. The distal part of the L2 loop is found in different conformations depending on the inhibitor bound and electron density often lacks for this part of the loop. There is also no sequence similarity in this loop amongst class I HDACs except for the binding rail region (Figure 1.11). In some crystal structures almost the entire L2 loop is missing electron density (PDB codes 3SFH, 1VKG) which is likely because of no favourable interactions with the L2 at the binding rail can be formed with the used inhibitors, i.e. ligands in these structures are not substrate like and can not form hydrogen bonds with Asp101 or exhibit π-stacking with Tyr100. The HDAC8 structure reported without a ligand bound shows no defined atoms in the deposited structure for residues 85 to 103 (PDB code 3F07 chain C).[127] Thus, all these studies suggest that the L2 loop is highly dynamic and the weaker binding to the binding rail results in a more dynamic loop.

The L1 loop spans residues 31 to 35 (Figure 1.10) and is also located in the vicinity of the entrance to the catalytic site. The L1 loop is directly above an internal cavity, which is believed to play a role in the catalytic mechanism

by functioning as an exit tunnel for the acetate product after catalysis[19] as described in more detail in Section 2.5. As for the L2 loop described above, the L1 loop is found in very different conformations in different crystal structures. Specifically, the C_α position of Lys33 in this loop deviates more than 5 Å when bound to different inhibitors (PDB codes 1T69 and 1T64), demonstrating its flexibility and ability to undergo large-scale motions.[126]

1.4.4. Functional Hot Spots in HDAC8

Correlation between HDAC8 activity and loop flexibility has been suggested previously by Somoza et al.,[125] Dowling et al.[127] and us,[1] as described later. In particular, it was shown that a zinc ion can bind to the distal part of the L2 loop and stabilising it (Figure 1.12 (a)), in terms of a more confined electron density in the crystal structures.[126,127,129] This proposed stabilisation down-regulates the activity 2-fold, thus suggesting that flexibility of this loop is important for function.[127] Moreover, a recent crystal structure and study of HDAC3[130] showed strong evidence that flexibility of both the L1 and L2 loop is imperative for catalytic function of HDAC3. HDAC3 and HDAC8 are both class I histone deacetylases and they have high sequence and structural similarity (41% sequence identity and a structure similarity Z-score of 55.8 for crystal structures 1T64 and 4A69), which indicate that flexibility of the L1 and L2 loops is also important for HDAC8 and possibly important for class I HDACs in general. A full sequence alignment of class I HDACs can be found in Figure A.1.

In addition to the regulation via the second zinc binding site, it was demonstrated that HDAC8 can be regulated via phosphorylation of Ser39,[131] which is adjunct to the L1 loop on the surface of HDAC8 facing the solvent. Phosphorylation of Ser39 down-regulates the enzyme. Moreover, HDAC8 has two potassium/sodium ion binding sites, which have different dissociation constants.[128,132] Each binding site was shown to have an effect on the activity, however, it is a

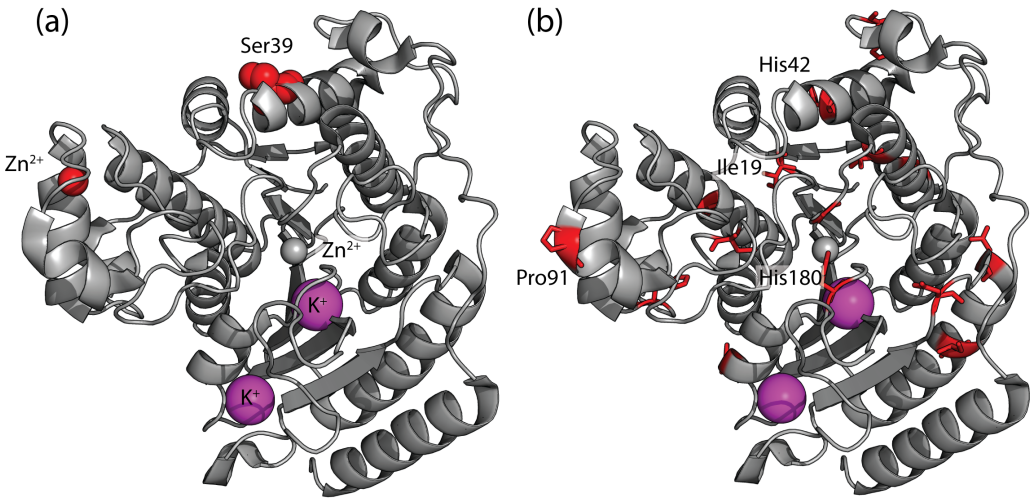


Figure 1.12.: Illustration of HDAC8 functional hot spots (PDB code 2V5W). (a) Regulation hot spots highlighted as spheres: Potassium ions in their respective binding sites are shown as magenta spheres, phosphorylation site Ser39 is shown in red and the second Zn^{2+} binding site is shown in red as well. (b) HDAC8 mutations identified in CdLS patients by Deardorff and co-workers[121] mapped on the crystal structure. Residues discussed in more detail in the main text are labelled.

current debate in the literature whether the ion that binds close to the active site up-regulates the activity of the hydrolase and ion binding to the second distal binding site down-regulates the activity or *vice versa*.[128,132–134]

On top of these regulatory mechanisms, the missense mutations linked to CdLS should be considered functional hot spots as they do change the activity of the enzyme. Here, especially mutations that can not readily be explained by clear structure-function interference are of interest from a functional point of view (Figure 1.12 (b)). Similarly, a mechanism for how the ion binding sites, phosphorylation site and loop dynamics tune the activity of HDAC8 had not been proposed. Hence, studying the atomistic dynamics of this enzymes could yield insight into the function of HDAC8 in terms of activity, regulation and pathogenicity via mis-function.

1.5. Research Hypotheses

The flexibility of HDAC8 seems to be an important feature of its function and hence needs to be studied in order to understand this enzyme. This would be in line with the emerging concept of a structure-dynamics-function relationship, which is essentially a paradigm shift from the structure-function relationship.[6,135,136] Deriving the function of macromolecules from their structure alone has been very successful even for large protein assemblies,[4] but there is evidence emerging that the dynamics or conformational sampling of a protein define or influence the mode of action of the protein.[6,17,137] Moreover, there are proteins, which do not have a defined globular native structure and are partly or fully unfolded, e.g. intrinsically disordered proteins, which still carry out specific functions.[138–141]

Based on the studies of HDAC8 my main research hypothesis is that dynamics do indeed correlate with the deacetylase activity of this hydrolase. While there are examples of proteins populating higher energy states[17] that could be related to function or dynamic altering mutations that change the catalytic rate of an enzyme,[137] implications of dynamics for e.g. ligand binding are possible as well. Hence, being important for the catalytic cycle of the enzyme.

As described earlier, understanding how HDAC8 works on an atomistic scale is vital to a) understand an important feature of the gene regulation machinery and b) aid disease treatment strategies via inhibition or regulation of HDACs. Hence, the main research aims can be summarised as follows:

- Study the structural dynamics of HDAC8, with focus on the L1 and L2 loops
- Identify structural and dynamic aspects or mechanisms that could interfere or promote function
- Investigate the role of the found mechanisms (e.g. regulation/catalysis/binding)

2. Results and Discussion

This chapter summarises the main results of my studies concerning HDAC8, which have in part been published in Ref. 1. First the MD simulations of HDAC8 are presented, developing a working hypothesis about the correlation of L1:L2 interactions and the function of HDAC8. This hypothesis is subsequently investigated further using single point mutations and NMR studies. Finally an investigation of the hypothesised acetate release tunnel is presented.

2.1. Molecular Dynamics Simulations

To study the dynamics of HDAC8 from a theoretical point of view I employed unbiased fully atomistic multi microsecond MD simulations using GROMACS and the Anton supercomputer as described in Section 5.6 and elsewhere.[1] HDAC8 was simulated in a free form, without any substrate or ligand, and complexed with the FDA approved inhibitor SAHA (Vorinostat), i.e. HDAC8:SAHA.

In all the simulations the overall structure is very stable, especially the hydrophobic core. However, as expected the loops of the enzyme are very dynamic, especially the L1 and L2 loop. Figure 2.1 shows the root mean square fluctuation (RMSF) at the region spanning the L1 and L2 comparing the flexibility of the loops in the $1.1\mu s$ trajectories of HDAC8 and HDAC8:SAHA. The L1 and L2 loops clearly experience different dynamics in the free form compared to the SAHA complexed one, the dynamics are attenuated in the SAHA complexed simulation.

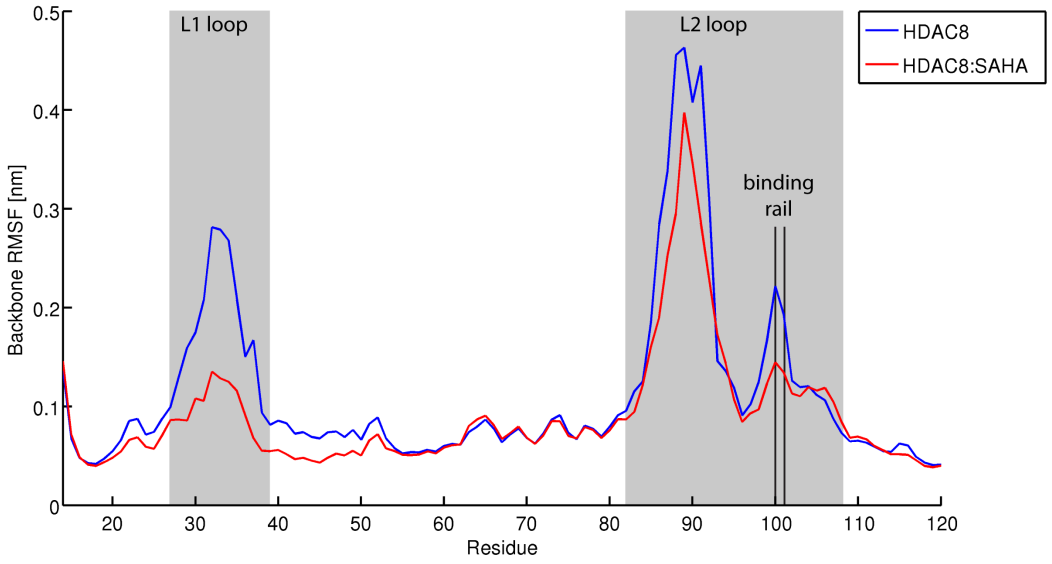


Figure 2.1.: Root mean square fluctuation (RMSF) of the backbone during the simulation of HDAC8 (blue) and HDAC8:SAHA (red). The region spanning the L1 and L2 loop (both highlighted with a grey background) is shown and the position of the binding rail residues are highlighted as well.

This led me to investigate these loop dynamics in more detail. As mentioned in Section 1.4.3 these loops are interesting from a structural and evolutionary perspective and comprise functional important residues, i.e. the binding rail (Figure 2.1). The MD simulations also showed that the loops can interact via electrostatic interactions. Therefore, the loop structures were analysed using clustering techniques and principal component analysis. The clustering (RMSD matrix in Figure A.2) showed that the loops sample three different states in the free HDAC8 simulation, and there seemed to be a correlation of the loop states and the dynamics of the binding rail. Hence, I analysed the trajectories accordingly to investigate possible effects of the described dynamic interactions on the functionally important residues of the binding rail.

2.1.1. Dynamic L1:L2 Loop Interactions

During the MD simulations the L1 and L2 loops of HDAC8 interconvert between different states and the conformational sampling observed compares well with crystal structure data (Video 1 - Appendix A.9). Interestingly, the binding rail shows two distinct conformations (coloured light green and dark green respectively, in Figure 2.2 (a)). These two states will be referred to as 'in' and 'out' state.

In the 'in' conformation Tyr100 and Asp101 build a rail towards the catalytic site as observed in crystal structures (Figure 1.9 (b) and Video 1 - Appendix A.9), which has been attributed to substrate binding and positioning.[126] On the contrary, in the 'out' conformation, the binding rail residues 100 and 101 are oriented away from the entrance to the catalytic site. The interconversion between 'in' and 'out' conformations takes place at Tyr100, as shown in the analysis of the backbone Φ angle (Figure 2.2 (b)). The Φ angle is larger than $-60°$ in the 'in' conformation, where for example the binding rail is mainly in the 'in' conformation between $800 - 1000$ ns.

Since the binding rail is very important for function,[126] mechanisms that interfere with its orientation are likely to correlate with the function of the enzyme. The earlier mentioned clustering connected the L1:L2 interaction with the binding rail conformation, this connection can also be made by observing the backbone behaviour of these regions during the simulation (Figure A.4). Furthermore, within the L2 loop a helix can form correlating with states found using clustering methods.

In total there are three features of the L1 and L2 dynamics that can be represented by three microkinetic processes: (1) the binding rail flips between 'in' and 'out' conformations, (2) L1:L2 interaction, dominated by electrostatic interaction between Lys33 of L1 and the triple Asp repeat of L2 (residues 87 to

89), and (3) formation of a short α-helix in the L2 loop (residues 93 to 97) as shown in Figure 2.2.

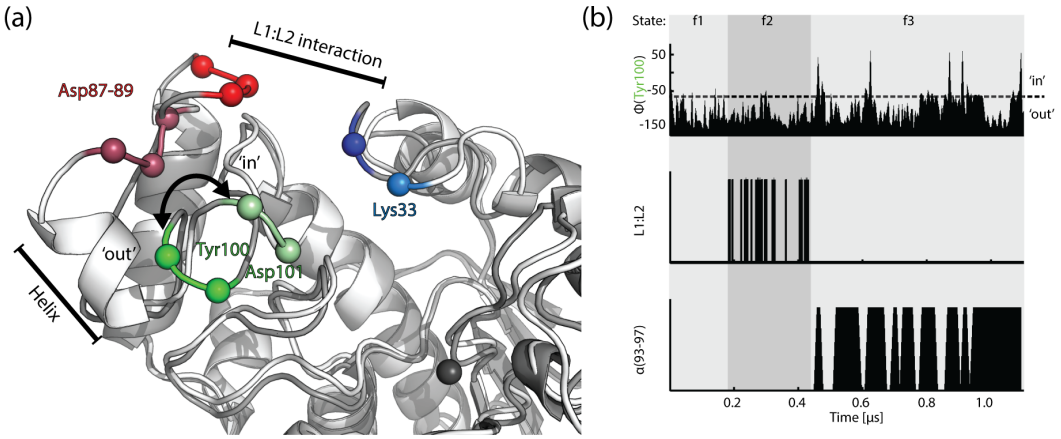


Figure 2.2.: (a) Ribbon representation of snapshots during the simulation showing the 'in' (pale colours) and 'out' (dark colours) conformation of the binding rail (green residues). Microkinetic processes including their localisation are annotated. (b) Microkinetic processes and states over the simulation time: 1) binding rail flips, measured via Φ of Tyr100 2) L1:L2 interaction between Lys33 and Asp87-89 measured using a Lys N_η to any of the six O_δ of the triple Asp repeat distance cut-off of 0.35 nm 3) presence of an α-helix at residues 93-97 as calculated by STRIDE. Figure reproduced from Ref. 1.

As mentioned earlier these processes seem to be correlated (Figure 2.2 (b)) and hence the trajectory can be divided into three consecutive states: f1, f2 and f3, which each show a characteristic pattern regarding the microkinetic processes. The binding rail initially starts in the 'in' conformation and flips to the 'out' conformation after a couple of nanoseconds. After the initial flipping out, the binding rail is only found in its 'out' conformation for the next 420 ns. This behaviour of the binding rail appears to be cooperative with conformational changes along the L2 loop from Tyr100 towards the triple Asp repeat and the formation of salt bridges between the triple Asp of the L2 loop and the Lys33 of the L1 loop, characteristic for the L1:L2 interaction. Thus, the binding rail is trapped in the 'out' state between 160 ns and 420 ns where the L1 and

L2 loop interact (Video 1 - Appendix A.9). The electrostatic interactions are not present in the subsequent f3 state, which allows the L2 loop to adopt conformations where the binding rail becomes loose and can flip between the 'in' and 'out' conformations. Thus, after 420 ns of simulation time, when reaching the f3 state, several 'in' to 'out' and 'out' to 'in' flipping events can be observed (Figure 2.4 (d)).

It seems that the binding rail behaviour can be steered by the presence or absence of the L1:L2 interactions. A number of transitions between the microkinetic states are observed, which suggests that the energy landscape between the states is shallow. Hence, a small perturbation to the energy landscape, such as point mutations or binding of small regulators and inhibitors could stabilise one of the states and thereby shift the populations, as has been shown for low-lying thermally excited states in other proteins.[17,142,143] To further investigate a possible correlation between the binding rail and the L1:L2 interaction the energy landscape of HDAC8 is perturbed by simulating the response to a small and flexible HDAC inhibitor, SAHA.

2.1.2. HDAC8:SAHA Simulation

The structure of the HDAC8:SAHA complex is shown in Figure 2.3. Flexible behaviour of the ligand and the loops at the entrance surface in the HDAC8:SAHA complex is expected, since crystallographic B-factors are large in these regions (Video 2 - Appendix A.11).

As for the HDAC8 simulation the trajectory has been decomposed into the three microkinetic processes (Figure 2.3 (b)), clearly revealing two distinct states, i1 and i2. In the i1 state, after the initial interaction of the binding rail with SAHA is lost, the aromatic ring of SAHA adopts a number of binding modes different from the crystal structure, although the hydroxamate group always remains bound to the catalytic zinc (Video 2 - Appendix A.11).

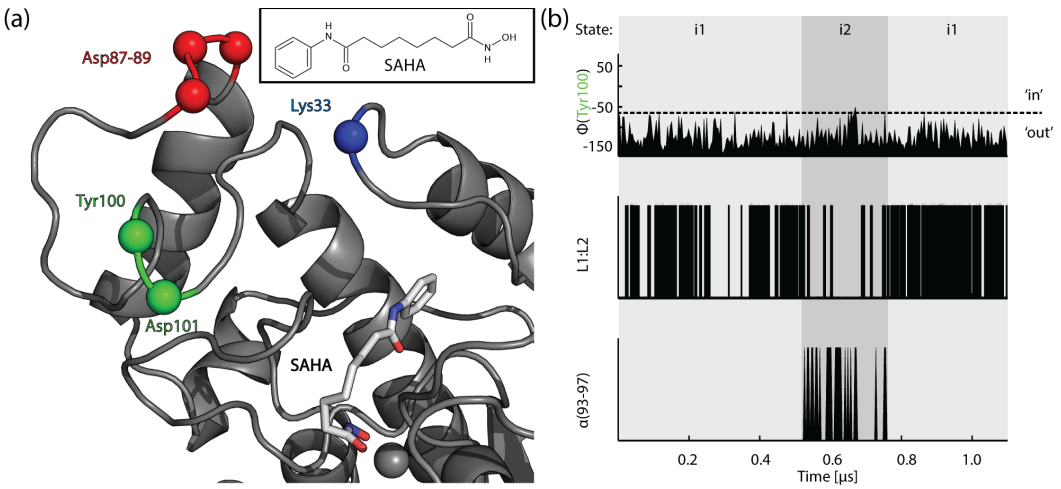


Figure 2.3.: States of the L1 and L2 loop conformations when interacting with SAHA (chemical structure shown in the insert). (a) HDAC8:SAHA complex snapshot of the simulation with SAHA (sticks) and HDAC8 (grey cartoon) where the binding rail is in its 'out' conformation. Positions of Lys33, Asp87-89, Tyr100 and Asp101 are illustrated with coloured spheres as in Figure 2.2. (b) Microkinetic processes over the simulation time as defined in Figure 2.2 1) Φ angle of Tyr100, indicating binding rail conformation 2) L1:L2 salt bridge presence between Lys33 and Asp87-89 3) presence of an α-helix at residues 93-97. Figure reproduced from Ref. 1.

Overall, in the i1 state SAHA can adopt a conformation where its aromatic cap group is buried in a pocket that becomes accessible in the vicinity of the entrance tunnel and the L1 loop. In this conformation, where SAHA is buried in the pocket at the L1 loop, Lys33 exhibits frequent electrostatic contacts with the triple Asp repeat of the distal part of the L2 loop, while the binding rail is in the 'out' conformation, preventing any contacts with the inhibitor. Thus, the frequent interactions of L1 and L2 loops prevent helix formation and keep the binding rail trapped in its 'out' conformation.

This is the same mechanism observed for the free form in the f2 state, where the distal part of the L2 loop interacts with the L1 loop. In the i2 state, which is reached after approximately 500 ns, the L1:L2 interactions are less frequent, allowing the formation of an α-helix in the L2 loop (residues 93 to 97). During

the i2 to i1 transition at 780 ns the L1:L2 interactions again increase, coinciding with an unravelling of the helix.

2.1.3. Conclusions of MD Simulations

Figure 2.4 shows a comparison of the loop structures during the simulations against published crystal structures, not the starting structure. The L1 loop clearly samples states that are also captured by different crystal structures where the loops show very different conformations (Figure 2.4 (a),(b),(e) and (f)), in which the C_α position deviates 0.5 nm. The crystal structures used for comparison are 2V5W (blue graphs), which is assumed to closely resemble a substrate bound state as described in Section 1.4.3, where the other structure (red graphs) is a structure of HDAC8 bound to two TSA inhibitor molecules. For the free HDAC8 simulation Figure 2.4 (c) and (d) highlight how closely such a binding conformation is sampled in the f3 state, which is not the case in the HDAC8:SAHA simulation.

Hence, the simulations seem to largely capture what has been known about the plasticity of the HDAC8 enzyme, but suggest a novel L1:L2 interaction. The L1:L2 interactions in the HDAC8:SAHA complex are much more pronounced than in free HDAC8, trapping the binding rail in its 'out' conformation throughout the simulation. Nevertheless, it is clear that the states sampled by the perturbed form (HDAC8:SAHA) and the free form of HDAC8 show many similarities, in particular, the f2 and the i1 state have many common features, such as the conformation of the L1 loop and frequent electrostatic interactions between Lys33 and the triple Asp repeat in the L2 loop. Hence, it can be concluded that the mechanism of L1:L2 interaction steering the binding rail behaviour is robust and can be triggered by small perturbations such as the binding of a small molecule.

Based on this it can be speculated that the L1:L2 interaction does have an effect on the activity of the enzyme, most likely the substrate binding or sta-

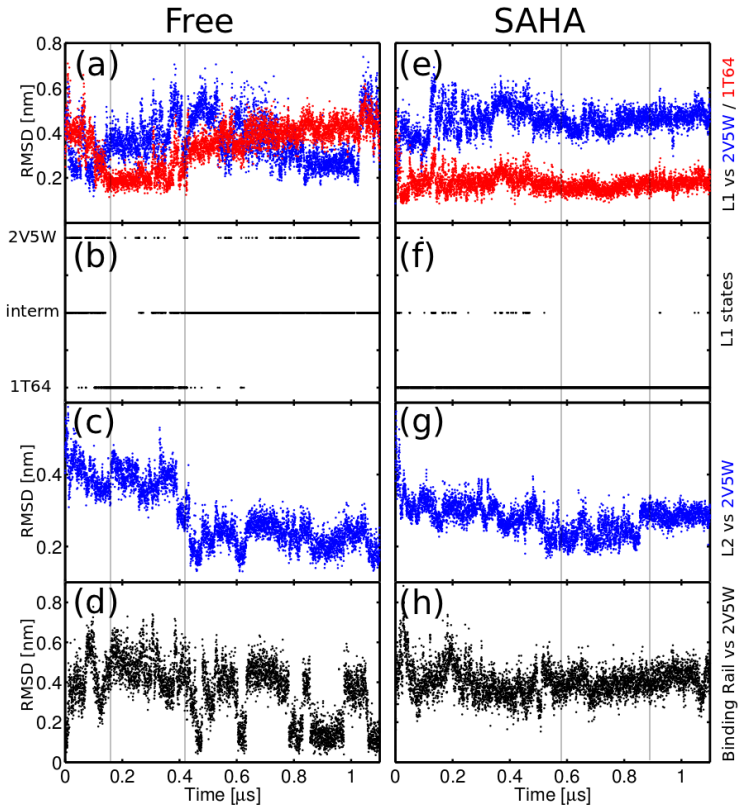


Figure 2.4.: Comparison of the L1 and L2 loop structures during the simulations with crystal structures. The left panel shows data for the free and the right panel data for the SAHA complexed form. (a) and (e): RMSD of the L1 loop against the binding state structure (2V5W in blue) and where the L1 loop is close to a conformation observed in 1T64 (red) after a least square fit of the rigid protein backbone. (b) and (f): State of the L1 loop clustered to the loop in the crystal structures 2V5W or 1T64 with a 0.26 nm cut-off and an intermediate mobile state. (c) and (g): RMSD of the whole L2 loop against the binding state structure of that loop found by Vannini et al.[126] in the crystal structure 2V5W after least square fit of the rigid protein backbone. (d) and (h): RMSD of the backbone of the binding rail residues Tyr100 and Asp101 after a least square fit of the rigid protein backbone. Figure reproduced from Ref. 1.

bilisation, as it influences the conformational state of the binding rail. Such a mechanism could explain the down regulation via the second Zn^{2+} at the L2

loop (Figure 1.12 (a))[127] or even the very subtle down-regulation of the CdLS mutant Pro91Leu.[123]

While this hypothesis seems viable, in the 1.1 μs simulations only a few transitions have been observed between the states. For a more significant and quantitative interpretation more sampling of the transition would be preferable. To enhance the sampling of the transitions I used metadynamics simulations and well-tempered metadynamics. However, these proved unsuccessful as using CVs such as the L1 to L2 loop distance and backbone angle of Tyr100 did not enhance sampling of the transitions consistently. It appears that there are transverse processes to the chosen collective variables that limit the sampling rate and lead to strong hysteresis, i.e. the choice of CVs was not sufficiently representative of the reaction coordinate.

A better choice of CVs might alleviate this. For example using native contacts as CVs in PTmetaD-WTE simulations could possibly enhance the sampling of the phase space sufficiently to yield a more accurate description of the free energy landscape the loop mechanism is experiencing.

As an alternative route to improve the sampling of the state transitions, I used the Anton supercomputer to carry out a much longer, unbiased, molecular dynamics simulation. While simulating free HDAC8 for 13.2 μs did not yield proportionally more sampling of the state transitions, the longer simulation confirms previous observation that L1:L2 interactions steer the binding rail backbone conformation (Figure 2.5) and several states can be seen where there are L1:L2 interactions present or absent.

Since these microkinetic processes steer a functionally important part of the protein, I hypothesised that the observed processes play a role in the catalytic cycle of HDAC8. It is unfortunately not feasible to simulate the whole catalytic mechanism including binding, cleaving of substrate and product egression. Even obtaining statistically robust measurements of the loop state transition is barely achievable, as shown by using Anton.

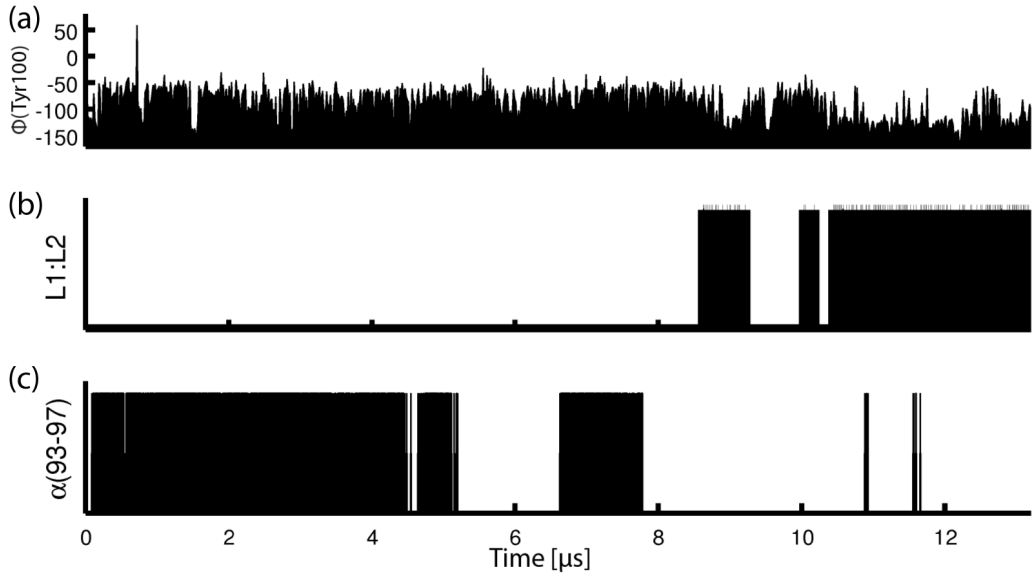


Figure 2.5.: The three microkinetic processes as a function of simulation time, as described in the text, for the 13.2 μs long Anton simulation of free HDAC8. Microkinetic states over simulation time; from the top a) binding rail flips around Φ of Tyr100; b) L1:L2 salt bridge presence between Lys33 and Asp87-89 as in Figure 2.2; c) presence of an α-helix at residues 93-97 as defined in Figure 2.2. In agreement with the shorter simulations, the Figure shows a correlation of the backbone angle Φ of Tyr100 on the interaction of the L1 and L2 loop and the probability for helix formation. Figure reproduced from Ref. 1.

Using an approach I termed potential energy perturbation calculation (PEPC), which is described in Section 5.4, with FoldX one can at least roughly estimate how mutations that interfere with the microkinetic processes would influence the state sampling (Figure 2.6). For the calculations the C_β positions of all residues were fixed to minimise the side chain reorientation by FoldX when mutating the residue. The PEPC result suggests that by mutating the important Lys33 to Glu the f2 state free energy is effectively increased by approximately 0.5 kcal mol^{-1} compared to the f1 and f3 state. Hence the f2 state should be 'de-populated' for this mutant.

Given the mentioned limitations and unsuccessful use of enhanced sampling

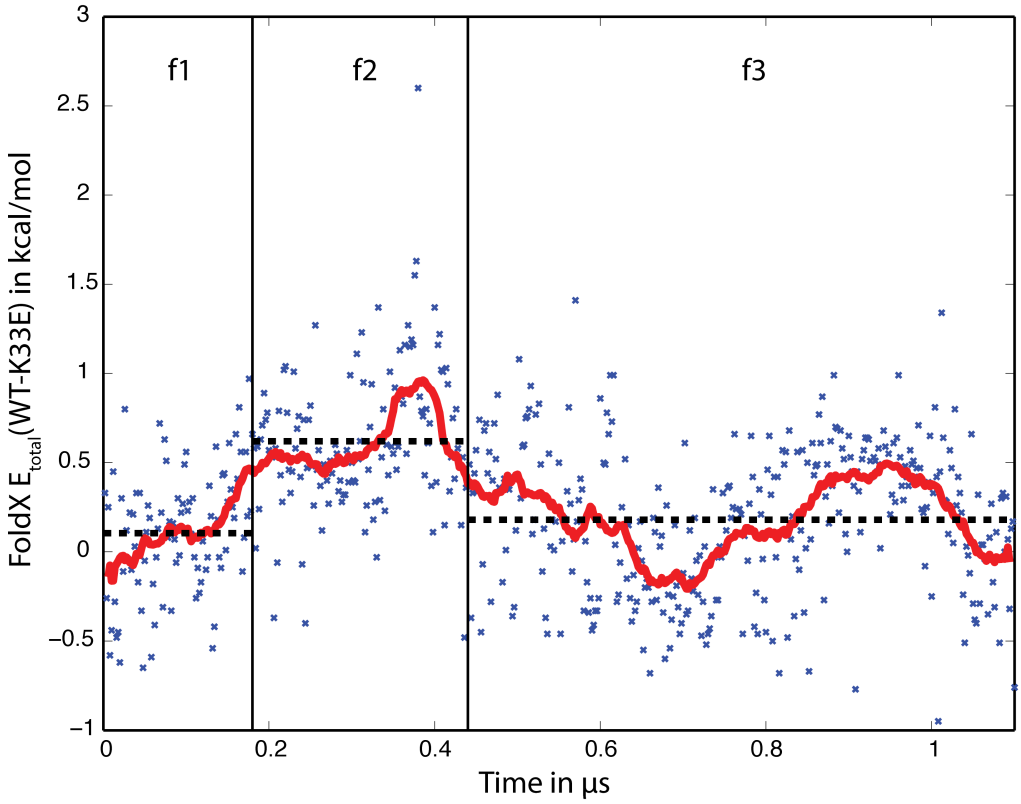


Figure 2.6.: Estimating the re-population of the states f1-f3 by introducing the point mutation Lys33Glu. FoldX was used to estimate the free energy difference in kcal mol^{-1} between the wt of HDAC8 and the mutant Lys33Glu for structures along the MD trajectory. Blue crosses show the data points from the FoldX calculations as the difference in the folding energy of mutant and wild type over time, the red line is a smoothed representation of the data and the dashed black lines indicate the average energy difference for each state. States f1-f3 are shown at the top of the Figure and separated by a black vertical line. The Figure clearly shows that in the case of Lys33Glu the f2 state, i.e. the state in which salt-bridges form between the L1 and L2 loop will be de-populated as the free energy difference is positive compared to the f1 and f3 state, the mean difference is 0.47 kcal mol^{-1}. Figure reproduced from Ref. 1.

schemes, only experimental evidence can substantiate this hypothesis. Moreover, force-field inaccuracies could lead to artefacts, which are however hard to

assess without converging simulations such as PTmetaD-WTE could possibly yield. Hence, I expressed HDAC8 mutants in which the L1:L2 interactions are altered and measured their activity, followed by structural and dynamical studies by NMR to test the working hypothesis derived from the MD simulations presented.

2.2. Studying the L1:L2 Interaction using Single Point Mutations

Validating hypotheses from theoretical work using key experiments is important to prove the relevance of made predictions. Especially in the presented case, where statistically robust sampling of transitions is hardly achievable, experiments have to be used to validate or refine the hypothesis. Even with recent advances in the development of force-fields, which can now accurately describe many dynamic processes,[31,144] force-field inaccuracies can not be neglected, highlighting the pressing need for a combination of computational and experimental work.

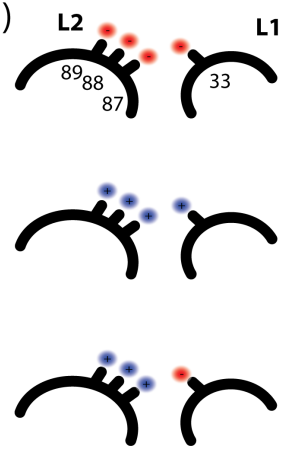
Based on MD simulations I hypothesised that the interactions between the L1 and L2 loops are capable of influencing the activity of HDAC8 via influencing the conformation of functionally important residues, i.e. the binding rail. To test this prediction I perturbed the interaction of the loops in a systematic mutational approach.

2.2.1. Loop Mutations

Wild type HDAC8 and mutants were expressed as described in Section 5.1.1 and elsewhere.[1,92] The electrostatic interaction between the L1 and L2 loop is characteristic for the L1:L2 interaction, which led me to design mutations that interfere with this microkinetic process by altering the charges on the loops: Lys33Glu, Asp87Arg/ Asp88Arg/ Asp89Arg, which is abbreviated as Asp87-89Arg, and a loop charge-swap mutant Lys33Glu/Asp87-89Arg (Figure 2.7 (a)).

The concept of this mutation scheme was to define a thermodynamic cycle. Mutants Lys33Glu and Asp87-89Arg should disrupt the electrostatic interaction of the loops by introducing only repelling electrostatic forces. The combination of both mutations, i.e. the Lys33Glu/Asp87-89Arg mutant, should then

(a)



(b)

Variant	Fluor. Assay [relative]	p53 Assay [relative]
WT	100%	100%
K33E	8% ± 2%	1.8% ± 0.2%
D87-89R	54% ± 7%	10.1% ± 0.5%
K33E D87-89R	15% ± 3%	11.5% ± 0.5%

Figure 2.7.: (a) Schematic representation of the studied mutations. From top to bottom: Lys33Glu, Asp87-89Arg, Lys33Glu/Asp87-89Arg. (b) Activity assay results of the studied mutants.

report on the dependence of both sites during the catalytic cycle. Hence, in the case where the mutation sites are completely independent the charge-swap mutant Lys33Glu/Asp87-89Arg should be simply the product of the other two mutations. However, this response does not yield information on how the site are depending on each other, in case they do. This approach is inspired by the concept of a thermodynamic cycle often used in protein folding studies.[145–149]

I first assessed the enzymatic activity of these mutants with a fluorogenic assay using Boc-Lys(Ac)-7-amino-4-methylcoumarin (MAL) as a substrate (Section 5.2).[150] In this assay the reaction takes place in a small volume for a fixed amount of time and is then quenched using an inhibitor. Subsequently the quenched reaction mix is developed using trypsin to cleave de-acetylated MAL product giving rise to an emission fluorescence wavelength change. Using the MAL substrate the wild type HDAC8 has a $\frac{k_{cat}}{K_m}$ of $38.34 \pm 3.6\,\mathrm{M^{-1}\,s^{-1}}$ (Figure 2.8). As summarised in Figure 2.7 (b), mutating Lys33 to Glu leaves the enzyme with only 8 % activity, whereas the mutation of the three Asp residues 87-89 results in 54.4 % residual activity. Interestingly, the charge-swap shows 14.9 % of wild type activity.

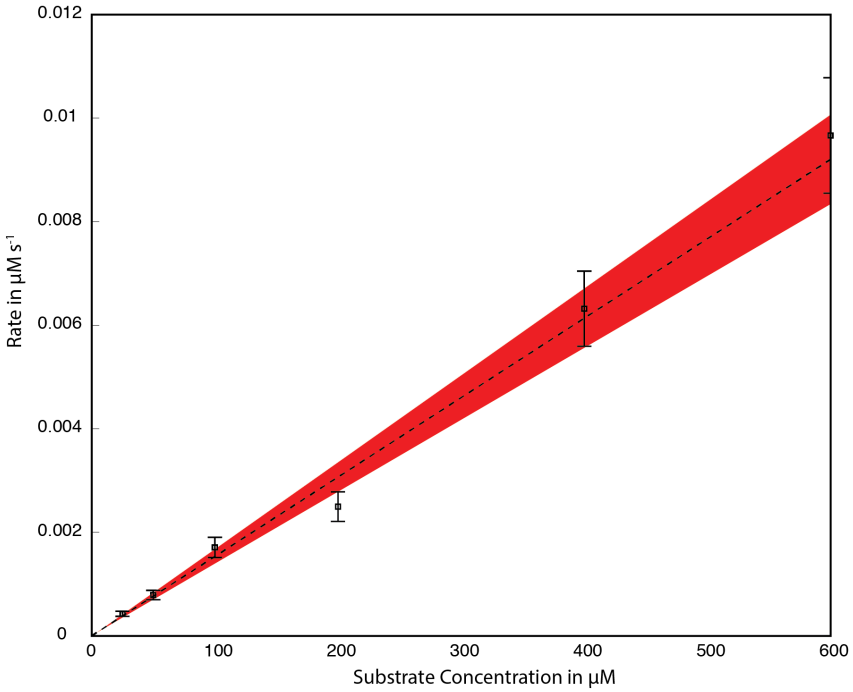


Figure 2.8.: Determination of $\frac{k_{cat}}{K_m}$ for the MAL assay using wild type HDAC8. Black squares show the measured rate with error bars showing the 2σ confidence interval obtained from deriving the rate constant from the fluorescence measurement. The black dashed line represents the best fit of a straight line to the data and the red area marks the 2σ confidence band from the χ^2 analysis. Figure reproduced from Ref. 1.

This suggests that interference with the electrostatic interaction of the loops has a detrimental effect on the enzymatic activity, although does not render the enzyme completely inactive. Furthermore, re-introduction of opposite charges on the loops in the charge-swap mutant results in an increased activity compared to the Lys33Glu mutant. However, the activity increase from Lys33Glu to Lys33Glu/Asp87-89Arg is small and does not exceed the activity of Asp87/89Arg.

It should be noted that Lys33 is not very far from the binding rail itself and mutation of this residue could directly interfere with substrate binding. Hence, the mutations are not solely influencing the electrostatic interaction of the

loops and some of the observed effects could be due to the change of charge and sterics. In general, the amino acids introduced (up to four) are of different sizes to those in the wild type enzyme, i.e. the surface is altered in shape and electrostatic potential, which could directly influence the binding of the substrate.

However, as can be seen from later described NMR experiments the influence of the e.g. the Lys33Glu mutation is very subtle overall and rather targeted (Section 2.4.3).

Moreover, the MAL substrate is by no means a perfect mimic of endogenous substrate since a fluorophore group is attached directly to the C-terminus of to the acetylated lysine, which like the substrate in the crystal structure 2V5W probably points directly towards Lys33 (Figure 1.9 (b)). It is therefore likely that a combination of a substrate specific effects and alteration of the L1:L2 interactions are leading to the observed changes in activity.

To address these issues I designed and employed a simple real-time NMR assay as described in Section 2.3 to test the activity of the HDAC8 and its mutants against a more natural substrate, i.e. a p53 derived peptide.[1] In this assay the wild type $\frac{k_{cat}}{K_m}$ is $289.41 \pm 5.6\,\mathrm{M^{-1}\,s^{-1}}$. As shown in Figure 2.7 (b), the Lys33Glu activity is minimal, with only about 2% of the wild type activity. The Asp87-89Arg mutant and the charge-swap mutant Lys33Glu-Asp87/89Arg show very similar $\frac{k_{cat}}{K_m}$ values, at approximately 10% of the wild type activity. Hence, in this assay the recovery of activity from the Lys33Glu mutation resulting from reintroducing attracting charges on the L1 and L2 loops is larger than in the fluorogenic assay and yields an activity that seems to exceed that of the Asp87-89Arg mutation alone.

Overall, the magnitude of activity change and recovery of activity is in the same range of that of other mechanisms that have been shown to down-regulate HDAC8. This includes down-regulation of HDAC8's activity by a second Zn^{2+} binding to the L2 loop and phosphorylation of Ser39.[127–129,131] At least the

latter has been associated with *in vivo* regulation of HDAC8, where the activity change resulting from the phosphorylation of Ser39 or the mimicking mutation Ser39Glu is 2 to 6-fold.[131] Moreover, the resulting enzymatic activities are also in the same ballpark as many of the CdLS mutants,[121,123,124] as mentioned in Section 1.4.4. Hence, the effect of the L1:L2 interaction on the activity observed should be considered significant in a physiological context.

These results show that a) there are substrate-specific contributions to the effects of the mutations, and b) the charge mutations in the L1 and L2 loops are not independent. While the data does not unambiguously show that loop interactions are indeed present and steer the binding rail, the data strongly suggests an interaction between the L1 and L2 loop that in some way influences the activity. Hence, the L1:L2 interaction should be considered an important component within the catalytic cycle of HDAC8. The explicit influence on the k_{on}, k_{off} and k_{cat} rates is hidden in the $\frac{k_{cat}}{K_m}$ ratio, but a recovery effect can consistently be observed when swapping the charges on both loops, supporting the working hypothesis that this mechanism has a direct influence on the catalytic cycle. The residues mutated are not conserved and a direct interaction between the L1 and L2 loop has not been observed before. Hence, the data supports the hypothesis derived from the molecular dynamics simulations and establishes a first atomistic mechanism that rationalises the dynamics of the loops in a structure-dynamics-function relationship.

2.2.2. Conclusions of Mutational Studies

The catalytic cycle of HDAC8 involves at least three steps, including binding of substrate, cleaving of the acetylated lysine, and release of the products. Characterisation of the interconversions between these states during catalysis is central to both, understanding the underlying mechanism of HDAC8 and facilitating development of novel inhibitors of HDACs. Indeed, there have been many com-

putational studies concerning the the catalytic cycle of the enzyme with special interest on the catalytic step[19,133,134,151] and the inhibition.[134,152–154]

The structure of an inactive mutant of HDAC8 bound to a cleavable substrate (PDB code 2V5W - compare Figure 1.9 (b)) has been previously determined and I will assume that this structure largely represents a substrate binding state. A particular feature of this structure is that the binding rail is in the 'in' conformation, forming hydrogen bonds with the substrate,[126] while there are no obvious contacts between the L1 loop and the triple Asp repeat of the L2 loop. Moreover, the L2 loop is structured in the binding state with defined atoms at both backbone and side-chains (Figure 1.9 (b)).

The simulations described in Section 2.1 suggest that the L1 loop conformations and dynamics are able to steer the structure of the L2 loop and thereby influence the conformational sampling of the binding rail. Only in the free form is there an effective sampling of the binding conformation of the L2 loop (Figure 2.2). In the f3 state, the binding rail frequently samples the 'in' conformation. When the distal part of L2 interacts with L1 it traps the binding rail in the 'out' state. Conversely, when there is no interaction with the L1 loop, the binding rail can dynamically sample the 'in' conformation, i.e. a binding state.

Using biochemical assays, I could show that the L1:L2 interaction is indeed important for the activity of the enzyme, further substantiating the interpretation of the simulations. I therefore hypothesise that the observed microkinetic processes, in particular the L1:L2 interaction, are able to tune the binding rail availability, hence tuning the enzymatic activity.

However, the observations are on a macroscopic level and as mentioned earlier introducing the mutations will have other effects than just perturbing the L1:L2 interaction. While 1D ^{1}H NMR spectra show the integrity of the fold of the mutants when compared to the wild type (Figure A.6) other subtle changes can not be observed with these experiments. Hence, it is necessary to study this in

more structural detail to a) further test the predictions and subsequent findings and b) understand which structural and dynamical changes are taking place. To achieve this I used advanced NMR techniques such as methyl TROSY type experiments and isoleucine labelled proteins as described in Section 2.4.

2.3. Real-Time HDAC8:p53-peptide Assay

HDAC8 activity is commonly assessed with fluorescence assays.[150,155] Mostly the commercially available Fluor-de-Lys assay (BIOMOL) or derivatives such as the MAL assay are used in the Hansen Lab.[150] Another example is the enzyme-coupled assay developed by Wolfson et al.[155] and a mass spectrometry (MS) based assay.[156] While the MS based assay is a viable approach and allows for the use of natural peptides this is, as the fluorescence based assays, limited by the amount of time domain information, i.e. each reaction has to be quenched before the measurement.

The fluorescence based assays rely on a fluorescence group (typically coumarins) directly C-terminal of the acetylated lysine. After the reaction the deacetylated lysine product is recognised by trypsin, which cleaves off the fluorescence group during the development step resulting in wavelength change of emission fluorescence.

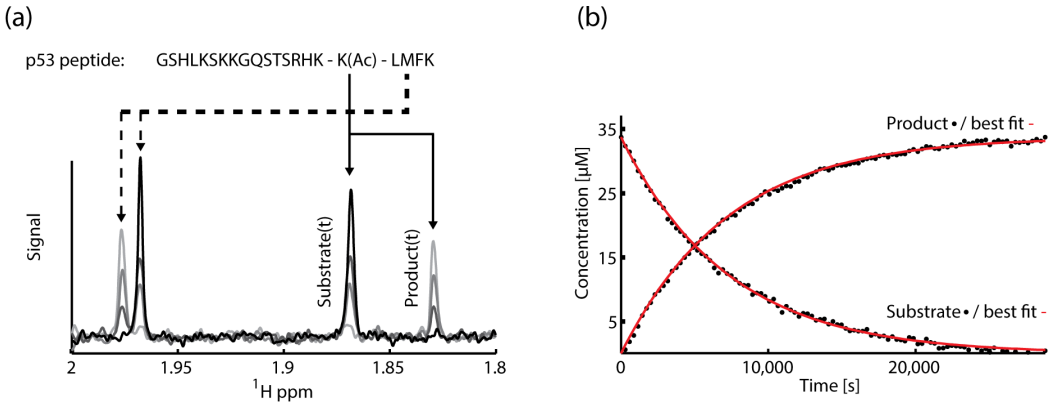


Figure 2.9.: HDAC8:p53-peptide assay. (a) 1D ^{1}H NMR spectra during a reaction, time elapsing from black to grey. (b) Integrated peaks of substrate and product scaled using a potassium acetate standard.

Hence, these assays face several drawbacks. The substrate used is not a very good mimic of endogenous substrates as it has a large fluorescence group attached. It has been noted by others that this could interfere with activity

measurements by e.g. artificially increasing the binding affinity of the substrate.[126] A fluorescence assay only allows one measurement per reaction as the reaction needs to be quenched. The solubility can also be an issue, especially for the MAL substrate, where the solubility of approx. 2 mM is lower than the K_m, which can only be estimated to be > 2.5 mM based on experience with the MAL assay, as the resulting reaction rates are in the $[S] \ll [K_m$ regime.

To alleviate some of these drawbacks I developed an NMR based assay to observe the reaction of HDAC8 against an acetylated peptide in real-time. This is based on the rather simple idea that in the NMR spectrometer only the peptide is observed with concentrations in the μM range and the enzyme stays 'invisible' at a concentration of 500 nM with its broad signal since its a large protein.

2.3.1. Setup

The assay is general and can in principle be used with any short peptide or potentially even larger substrates when combined with specific labelling and ligation schemes. Initially, the assay uses a 21 amino acid long peptide based on the flexible tail of the tumour suppressor p53, where the acetylated Lys corresponds to position 382 in the tumour suppressor sequence (see top of Figure 2.9 (a)). HDAC8 has been found to be active against a similar but shorter peptide[157] and there is a study identifying p53 as a possible endogenous substrate of HDAC8.[95]

The peptide is dissolved in a buffer similar to the one used in the Fluor-de-Lys or MAL assay, omitting the BSA in order reduce crowding of the NMR spectra (Section 5.2). The peptide concentration is measured using the NMR spectrometer by comparison to a potassium acetate standard. The sample is removed from the magnet and the enzyme is added. After mixing the sample is inserted into the spectrometer and a series of 1D ^{1}H NMR spectra is recorded

resulting in a small dead time of $3-5$ min for the first point, which has to be taken care of at data post-processing.

Important peptide resonances that report on the substrate-to-product conversion were assigned by comparing non-acetylated and acetylated peptide spectra (see Figure 2.9 (a) for assignment). Figure 2.9 (a) shows a typical ^{1}H spectra of a HDAC8 wt reaction, where the black spectrum is the first acquired spectrum and selected following spectra are coloured more and more grey as time elapses. The resonances observed are sharp as the peptide is very flexible and thus should behave similar to the p53 tail, which is disordered in solution.[139,158,159]

2.3.2. Data Analysis

To obtain the product and substrate concentrations over time the peaks are integrated, assuming that the peak volume is proportional to the concentration. This is more stable than using only the peak height due to the sharpness of the peaks and possible shimming problems during measurement time, which can slightly broaden the spectrum at later times, as can be seen in Figure 2.9 (a). To fit the acquired data (Figure 2.9 (a)) and obtain kinetic rates such as $\frac{k_{cat}}{K_m}$ I wrote a set of MATLAB routines. In the routines the χ^2 of a numerical solution of the Michaelis-Menten ordinary differential equations (ODEs) (Section 5.2.1) is first calculated for a grid of k_{cat} and K_m values (Figure 2.10 (a)). In the case of a single measurement or similar substrate concentrations in repeats, the χ^2 surface will have a minimum error path or line (black line Figure 2.10 (a)). This path is equal to $\frac{k_{cat}}{K_m}$. To improve the fit of $\frac{k_{cat}}{K_m}$ the perpendicular error to the best solution is calculated, using much finer spaced points (white line). The standard deviation σ can then be directly extracted from the last calculated reduced χ^2, i.e. $\sigma = \frac{k_{cat}}{K_m}(\chi^2_{min} \pm 1) \mp \frac{k_{cat}}{K_m}(\chi^2_{min})$ (Figure 2.10 (a)).
It is not strictly necessary to use a perpendicular slice to fit the error correctly as the solution for $\chi^2_{min} + 1$ would yield the same result in the case where only

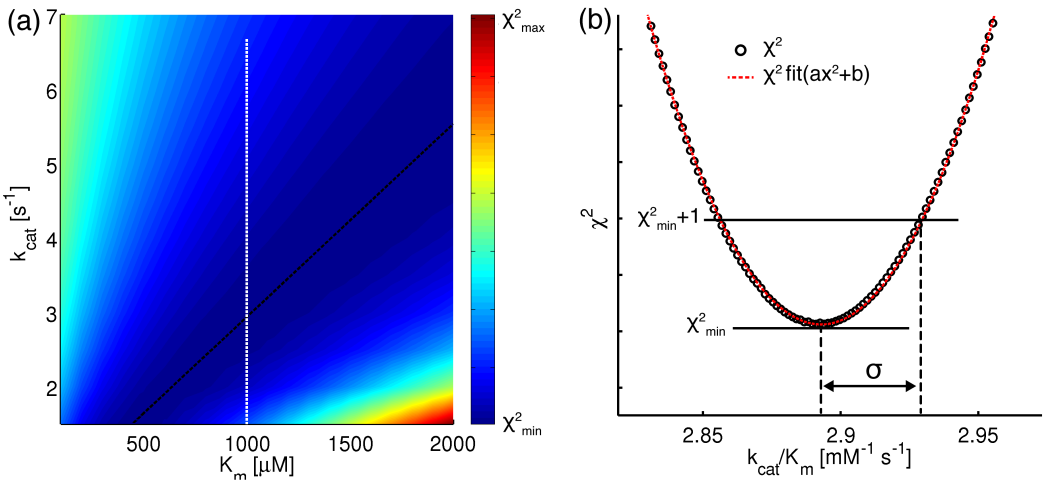


Figure 2.10.: Data analysis of the HDAC8:p53-peptide assay. (a) χ^2 error surface of the chosen k_{cat} and K_m values. Red colour marks a high, blue colour a low error. The black dashed line signifies the first estimate of the minimum error path. The white dashed line represents the perpendicular line that is used to refine the fit and calculate the statistical error σ. It should be noted that ratio of the axis are not 1:1. (b) χ^2 along the perpendicular slice, i.e. the white dashed line from panel (a). Black shows the χ^2 values and the red dashed line a fit using simple quadratic function $ax^2 + b$. Calculation of the statistical error σ is highlighted within the graph. The data presented stems from a current collaboration.

$\frac{k_{cat}}{K_m}$ is considered. In the case where these constants can be fitted individually, i.e. multiple different substrate concentrations, the minimum path approach would not work as described. However, the code can be easily changed to search for a minimum in the grid and from there the errors for the constants can be determined individually.

The error in calculating χ^2 is the standard deviation of the NMR signal, i.e. the noise level, which due to the sharp peaks is found to be around 5 % for μM peptide concentrations on the used Bruker Avance 500 NMR spectrometer (Section 5.3).

2.3.3. Cross-Validation

Since the buffer conditions and volume for the NMR assay are different the NMR assay was also run using MAL substrate to relate it to the fluorogenic assays. Figure 2.11 depicts the NMR assay when used with MAL. Unfortunately the substrate peak is broad and overlaps with acetate product peak making it more difficult to directly extract the kinetic data via peak integration. Hence, I used one-dimensional peak fitting in FUDA,[160] written by D Flemming Hansen (last accessed at: http://www.biochem.ucl.ac.uk/hansen/fuda/), to extract the peak integrals at each time point.

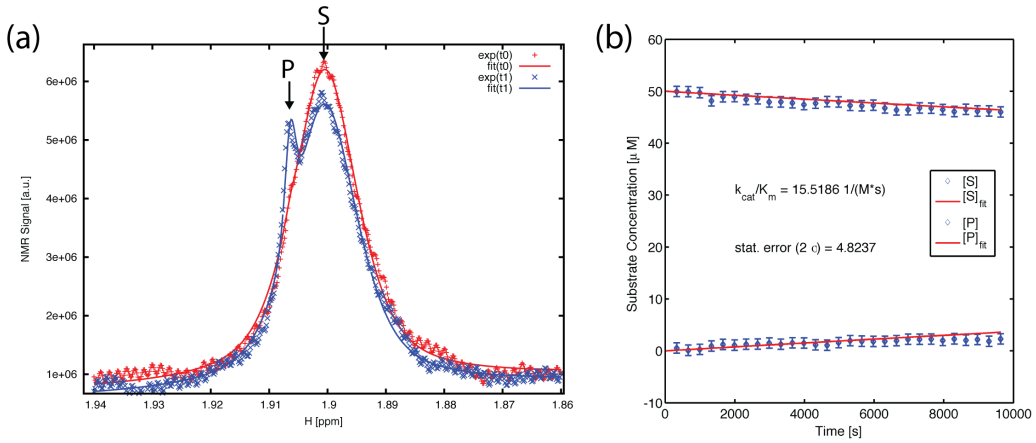


Figure 2.11.: Illustration of NMR assay using MAL substrate. (a) Part of the 1D ^{1}H NMR spectrum showing the broad substrate peak at the start of the assay (red) and the sharp acetate product peak coming up a time t1=2h in the experiment (blue). Points (+/x) show experimental data and lines show fits from FUDA. Product (P) and substrate (S) peaks are highlighted. (b) Kinetic data extracted using FUDA (blue points) and fits of the substrate and product peaks (red) obtained via global fitting of the Michaelis-Menten ODEs.

Analysis has been done as described earlier yielding a $\frac{k_{cat}}{K_m}$ of $\approx 15.5 \pm 5\, M^{-1}\, s^{-1}$, which is about half the value found for this substrate using the fluorescence assay in Section 2.2. However, this difference can be caused by factors mentioned earlier, such as the slightly different buffer conditions. Moreover, since

the peak is broad it might be that the fluorescence compound is not entirely in solution or sticking to the glass of the NMR tube (in the fluorescence assay a plastic 96-well plate is used). Since the activity observed is in the same ball-park as in the fluorescence assay it gives at least a reasonable reference for the NMR assay experiments.

2.3.4. Advantages and Limitations

The biggest limitations of the described assay are the money and time requirements. The peptide used is rather expensive and the spectrometer time can be quite rare for such experiments at a given institution. Another potential drawback is the time scale of the reaction, e.g. the reaction is already completed in the dead time. This means that the reaction is completed before the first measurement can be taken. Finally, spectral overlap can interfere with the assay.

However, many of these drawbacks can be circumvented to some degree. The cost can be reduced by using shorter peptides or producing the acetylated peptide by purifying the product peptide and acetylating it *in vitro*. In cases where the time scale of the reaction is not suitable for NMR detection the amount of substrate or enzyme can be tuned to get into the right regime. Even spectral overlap can be dealt with in different ways; as in the case of the MAL substrate some sophisticated fitting to deconvolute overlapping peaks can help or, in the case of using larger proteins as substrates, spectral overlap could be circumvented with clever isotope labelling and possibly natural ligation.

In this scenario specific amino acid labelling could be employed or isotope labelled acetate for the acetylated lysine group, e.g. using a deuterated background and proton labelled acetate. In a specific scenario where the modified part is a disordered tail, such as histone tails or the p53 C-terminal tail, this tail could be ligated to the deuterated protein. Of course, using multi-dimensional NMR techniques is possible, however this would in most cases reduce the time-

resolution as more scans would be required for the sampling of the indirect dimension.

The distinct advantage of this assay is that it is capable of using natural substrates during the reaction and therefore allows a non-invasive observation of the real-time kinetics of the system. While I used this approach in a rather simple form, the assay comes to its own by studying substrates and complexes that have different PTMs or PTM erasers (e.g. (de-)methylation and (de-)acetylation) to elucidate if and how the removal of these PTMs are inter-dependent. Let S_{ab} be the substrate containing both modifications a and b and let E_a and E_b be the enzymes that remove these modifications respectively. The substrate has to be 'processed' by both enzymes to become the product P that has none of the two modifications left. Consider that activity of E_a against S_{ab} is very high producing S_b very quickly, while the $[E_b] + [S_{ab}] \rightarrow [E_b S_{ab}] \rightarrow [E_b] + [S_a]$ process is very slow. For these intermediate products however, the activity of the PTM eraser can be different, e.g. $[E_b] + [S_b] \rightarrow [E_b S_b] \rightarrow [E_b] + [P]$ being much faster. Such inter-dependencies should be directly observable in this assay and can be analysed by a simple modifications of the fitted ODEs.

Indeed, there are complexes that comprise HDACs and demethytelases and this assay should be capable of measuring any inter-dependencies of these two PTM erasers. I am now actively working on such a system in collaboration with a different laboratory.

2.4. Investigating the L1 and L2 Structure and Dynamics using NMR

To test my working hypothesis regarding the dynamic L1:L2 loop interactions further I applied advanced NMR techniques in combination with amino acid specific isotope labelling schemes to elucidate the loop dynamics and their behaviour when altering the L1:L2 interaction via mutations.

Specifically, I used an isoleucine $^{13}CH_2D$ methyl isotopomer labelling scheme to investigate the dynamics of these side-chains via deuterium relaxation. In this regard I investigated whether published NMR pulse schemes could be further optimised. Subsequently the wt data are compared to the mutant data.

2.4.1. NMR Labelling Scheme

Despite exhaustive efforts in the Hansen lab, mainly by Nicolas Werbeck, we still lack a backbone $^{1}H-^{15}N$ resonance assignment of HDAC8, which would have enabled me to study backbone dynamics of all residues of the enzyme. However, using specific amino acid labelling all visible isoleucine $^{1}H_{\delta1}-^{13}C_{\delta1}$ methyl resonances could be assigned in $^{1}H-^{13}C$ correlation experiments. This was achieved by Nicolas Werbeck and two students (Havva Yalinka and Ruth Dingle née Pritchard) using Ile to Val mutations. HDAC8 has 25 isoleucines and thus yield reporters on most parts of the structure of HDAC8 (Figure 2.12 (b)). Most notably for my investigations are Ile34 and Ile94, which are located next the important Lys33 and in the part of the L2 loop that can form a helix respectively (Figure 2.12 (b)).

To study the dynamics I aimed to measure relaxation rates within the $C_{\delta1}$ methyl group to yield Lipari Szabo order parameters via mapping of the spectral density functions. There are different isotopomers of methyl groups available for NMR studies, i.e. $^{13}CH_3$, $^{13}CH_2D$, $^{13}CHD_2$ and $^{13}CD_3$. Each of them comes with some unique advantages that have been discussed in detail else-

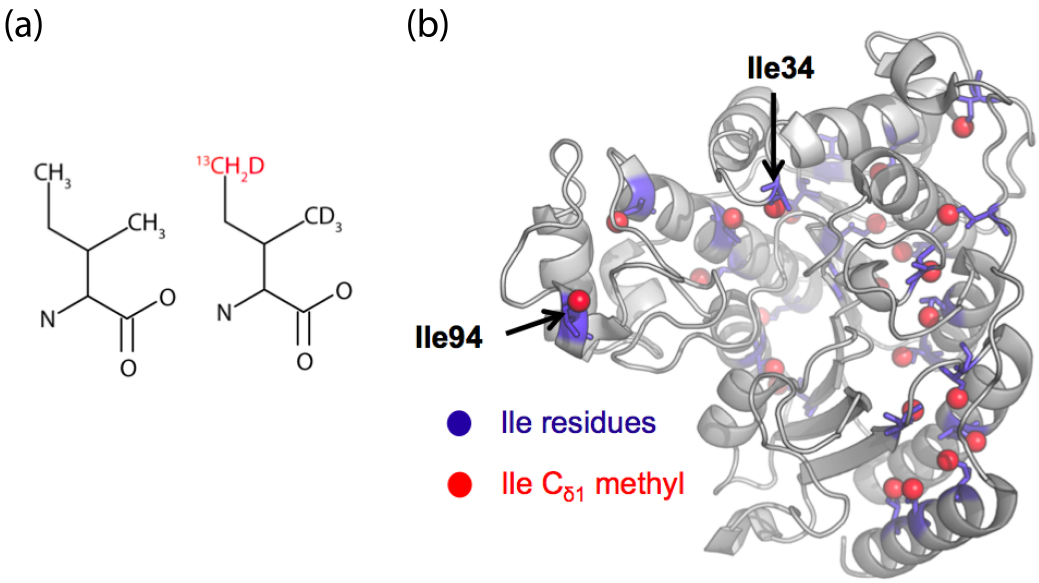


Figure 2.12.: (a) Structure of isoleucine with a CH_3 isotopomer (left) and $^{13}CH_2D$ isotopomer (right), highlighting the methyl group in red. (b) Cartoon representation of HDAC8 (PDB: 2V5W), highlighting isoleucine residues in blue and the $C_{\delta1}$ position in red. Position of Ile34 and Ile94 are marked with black arrows.

where.[71,74,82,161–163] I chose to use the $^{13}CH_2D$ (Figure 2.12 (a)) isotopomer for Ile labelling as this probe a) yields a better signal for medium and large proteins by using a $^{13}CH_2D$ TROSY type detection (Section 2.4.2) and b) due to the self-consistency of the five measurable deuterium relaxation rates of the spin system when fitting LS2 order parameters as introduced in Section 1.3.3. The labelling protocol for this scheme is described in Section 5.1.1.

2.4.2. $^{13}CH_2D$ NMR

Proteins were isotope labelled using an α-keto-buteric acid precursor with otherwise only deuterated media as detailed in Section 5.1.1 to yield Ile($^{13}C_{\delta1}H_2D$ only) labelled samples, i.e. only 50 (2 hydrogens per $C_{\delta1}$ methyl group) hydrogens per HDAC8 protein and 25 ^{13}C spins. Studying HDAC8 using this labelling scheme requires non-standard methods, which are detailed later in this

section. Moreover, since the used precursor yields Ile residues where only the $C_{\delta 1}$ carbon is labelled (i.e. $^{13}C_{\delta 1}$) I also investigated the possibility to design pulse sequences that take advantage of the fact that there is no decoupling for neighbouring ^{13}C spins needed as is the case for uniformly ^{13}C labelled proteins.[82,164]

$^{13}CH_2D$ Energy Levels

Figure 2.13 shows the energy levels, transitions and transverse relaxation rates of a $^{13}CH_2$ spin system, which represents a rapidly rotating $^{13}CH_2D$ methyl group when the deuteron is treated as 'silent'. Assuming that the methyl group is rotating very fast around the three-fold symmetry axis (on an NMR time scale, i.e. much faster than $1/\omega$, where the ω is the Larmor frequency) is important as only in this case the hydrogen nuclei spins can be treated as magnetically equal. Rapid rotation (approximately 20 ps)[82] is generally the case for Ile $C_{\delta 1}$ methyl groups and other methyl groups in proteins at ambient temperature and is therefore considered here.

It is noteworthy that the transverse relaxation of the $|\alpha\rangle|2\rangle \leftrightarrow |\beta\rangle|2\rangle$ transition $R^s_{2,C}$ is very slow as the intra-methyl dipolar relaxation contribution is to a very good approximation 0 considering only the $J(0)$ terms, which are the dominant terms in the macromolecular limit ($\omega_C \tau_C \ll 1$, where ω_C is the ^{13}C Larmor frequency and t_C is the overall tumbling time of the molecule).[71,82,165] The $|\alpha\rangle|2\rangle \leftrightarrow |\beta\rangle|2\rangle$ is the central transition of the carbon 1:2:1 triplet, as can be seen in the energy diagram (Figure 2.13). This line does not evolve under scalar coupling.[166]

These advantages can be exploited in methyl TROSY experiments as discussed later leading to sharp peaks due to the slow relaxation. In general methyl TROSY experiments, i.e. HMQC, have been shown to be superior to HSQC type sequences. However, for measuring fast relaxations of deuterium spins I investigated whether it would be favourable to use an HSQC sequence with

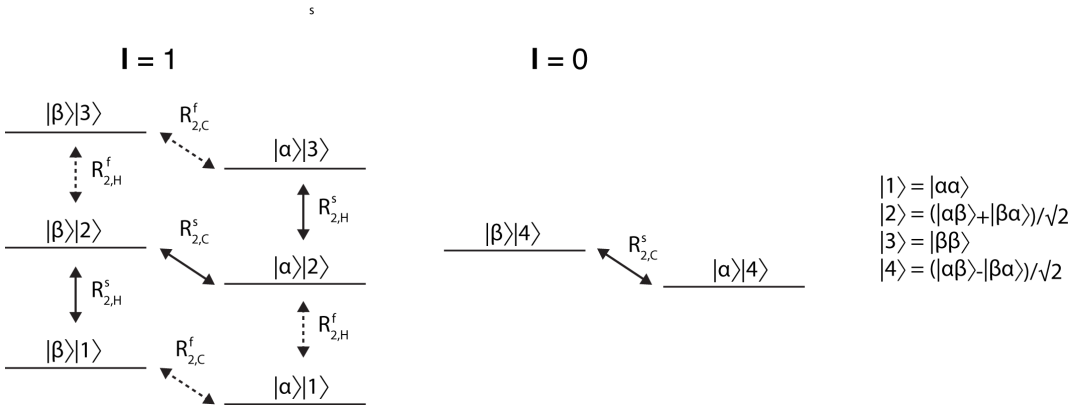


Figure 2.13.: Energy levels of a rapidly rotating $^{13}CH_2$[D] spin system. The deuteron of the $^{13}CH_2D$ methyl group is treated as a 'silent' spin for simplicity. I is the total spin quantum number of the system. Slow (fast) relaxing ^{13}C/H transitions are shown with solid (dashed) arrows and labelled with $R^s_{2,X}$ ($R^f_{2,X}$) where X denotes the nucleus transition, i.e. ^1H or ^{13}C. Spin states are denotes as $|j\rangle\,|k\rangle$, where $j = \{\alpha, \beta\}$ is the ^{13}C spin state and $k = \{1 - 4\}$ is the wave function of the equivalent methyl ^1H spins shown on the right.

the ^{13}C labelling scheme, i.e. only isoleucine $^{13}C_{\delta 1}$. The previously established HSQC experiments[74,163,164] use a constant-time element and shaped pulses to decouple $^{13}C_{\delta 1}$ from other carbons, e.g. $^{13}C_{\gamma 1}$ or $^{13}C_{\alpha}$, which in this case are NMR silent since these atoms are ^{12}C isotopes.

Removing these blocks from the sequence shortens the acquisition sequence resulting in a better signal due to less signal loss via relaxation. In an HSQC this can be essential as transitions are usually mixed during acquisition, which results in faster relaxation compared to TROSY type experiments as described in Section 2.4.2. In general, HSQC type sequences are taking advantage of all available magnetisation for the cost of mixed transitions and faster relaxation, where for TROSY, especially the $^{13}CH_2D$ TROSY, specific transitions are selected where only part of the magnetisation can be transferred. In the presented case ($^{13}CH_2D$) this leads to theoretical loss by a factor $2/\sqrt{2}$ for

the TROSY compared to the HSQC experiment.[82] Hence, an optimised HSQC scheme might be superior to a TROSY type experiment.

HSQC Deuterium Relaxation Sequence

To design a pulse sequence that was capable of measuring the longitudinal and transverse relaxation rates of the deuteron in the $^{13}CH_2D$ methyl group $R_{1,D}$ and $R_{1\rho,D}$ I started with a HSQC sequence as in Ref. 164. To improve the sequence part of the decoupling and the constant time element was removed. Figure 2.14 shows the final HSQC pulse sequence. The evolution of the magnetisation will be described in brief and the normalisation factors of the coherences are omitted: Transverse proton magnetisation $-H_y$ is produced by the first $\frac{\pi}{2}$ proton pulse and carbon proton scalar coupling is allowed to evolve during the INEPT part (Section 1.3.1), yielding H_zC_z magnetisation at point (a). Methyl proton magnetisation is represented by the operator $H_\alpha = H_\alpha^1 + H_\alpha^2$, $\alpha \in x, y, z$, i.e. the combination of both proton spins of the $^{13}CH_2D$ methyl isotopomer. The delay τ_a is chosen such that $\tau_a = \frac{1}{4J_{CH}}$, where J_{CH} is the one bond carbon proton coupling of the $^{13}CH_2D$ group. In practice this delay (and all other INEPT delays) are chosen a bit shorter to account for magnetisation loss due to relaxation, hence it is optimised during the experimental setup.

The following carbon $\frac{\pi}{2}$ pulse generates $-H_zC_x$ magnetisation and scalar coupling of the carbon and deuterium is allowed to evolve during another INEPT. Here, the delay is set to $\tau_d = \frac{1}{8J_{CD}}$ since deuterium has a nuclear spin of 1. This yields a mixture of coherences at point (b). As described by Muhandiram

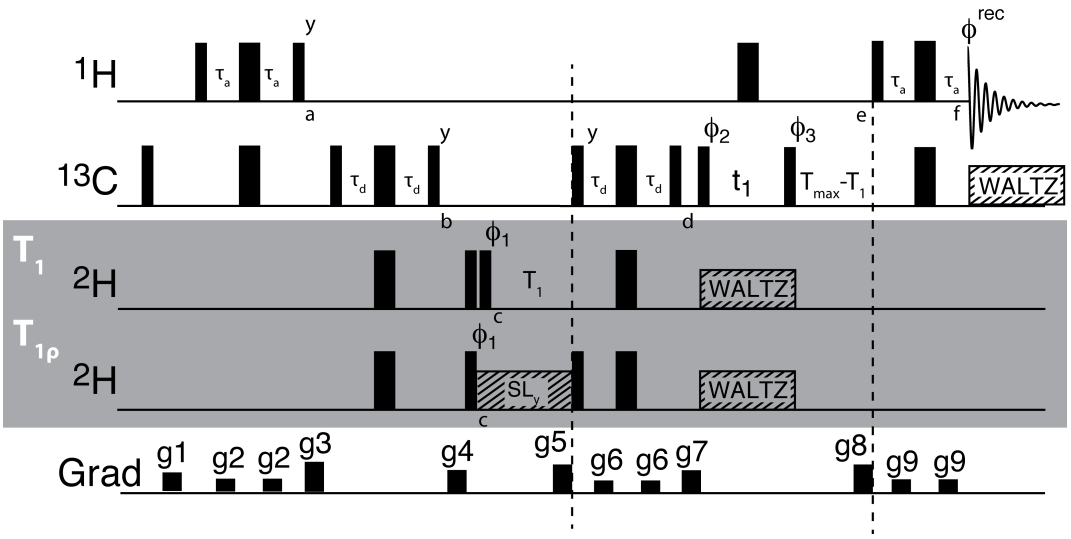


Figure 2.14.: HSQC pulse sequences to measure longitudinal relaxation rates/times ($R_{1,D}/T_{1,D}$) and transverse relaxation rates/times via ($R_{1\rho,D}/T_{1\rho,D}$) of a deuteron in a $^{13}CH_2D$ spin system. Narrow (wide) blocks represent $\frac{\pi}{2}$ (π) pulses around x unless otherwise noted. Points (a-f) are guides for the magnetisation evolution discussed in the main text. Full details for delays and phase cycling can be found in Section 5.3.

et al.[164] the triplet of the ^{13}CD spin system (i.e. transverse ^{13}C magnetisation) can be represented by the following carbon coherences:

$$C_x\left(1-D_z^2\right), \text{center line} \tag{2.1}$$

$$0.5\,C_x\left(D_z^2+D_z\right), \text{outer line} \tag{2.2}$$

$$0.5\,C_x\left(D_z^2-D_z\right), \text{inner line} \tag{2.3}$$

This description can be easily followed by realising that these coherences are eigenfunctions of the scalar coupling Hamiltonian $H^J = 2\pi J_{CD} C_z D_z$, which can be proven by calculating the commutator of these coherences with $2\pi J_{CD} C_z D_z$ (Appendix A.2). The $^{13}C-D$ scalar coupling evolution can then be written as:[164]

$$C_x \rightarrow C_x\left(1-D_z^2\right) + C_z D_z^2 \cos\left(4\pi J_{CD}\tau_d\right) + C_y D_z \sin\left(4\pi J_{CD}\tau_d\right), \tag{2.4}$$

which immediately shows that in order to remove any D_z^2 components from the spectra a phase cycling is required. Furthermore, formula 2.4 shows that using a delay of $\tau_d = \frac{1}{8J_{CD}}$ allows two thirds of the magnetisation to transfer in this ^{13}CD spin system since $C_x D_z^2$ represents the sum of the inner and outer line of the 1:1:1 triplet.

For the T_1 experiment, i.e. to measure the longitudinal relaxation, two $\frac{\pi}{2}$ pulses are given, where the latter one is phase cycled $\phi_1 = +x, -x$ generating $\pm H_z C_z D_z$ magnetisation, hence changing the sign of the signal, which is taken care of by changing the phase of the receiver ϕ_{rec} in concert with ϕ_1. In contrast to the $C_y D_z$ term (Formula 2.4) the D_z^2 terms will not change sign and are hence removed via the phase cycling. This phase cycling step is essential to ensure that the magnetisation present during the relaxation delay T_1 is proportional to D_z.

The relaxation of the $H_z C_z D_z$ magnetisation (point (c)) during the subsequent relaxation delay T_1 is dominated by the fast deuterium relaxation due to the quadrupolar interaction. While there is a contribution from the H_z and C_z relaxation, this can be subtracted 'on the fly' immediately before (e) by employing a $T_{max} - T_1$ delay while only $H_z C_z$ magnetisation is present, i.e. the contribution of the H_z and C_z relaxation is constant for each increment of the T_1 delay measured. After the T_1 delay $H_z C_x D_z$ is generated by a carbon $\left(\frac{\pi}{2}\right)_y$ pulse and with another INEPT transferred to $H_z C_z$ magnetisation at point (d).

For the $T_{1\rho}$ experiment the first $\frac{\pi}{2}$ deuterium pulse is omitted to yield $\pm H_z C_z D_y$ magnetisation at point (c) which is subsequently spin-locked by the application of the SL_y pulse for time T_1. D_z^2 terms establish double quantum coherences, which are filtered out using gradient selection. To transfer the magnetisation from the deuterium back to the carbon nucleus $H_z C_x D_z$ magnetisation is generated by simultaneously applying a carbon $\left(\frac{\pi}{2}\right)_y$ and a deuterium $\frac{\pi}{2}$ pulse. As for the T_1 experiment magnetisation is transferred to $H_z C_z$ (point (d)) via an INEPT sequence.

After point (d) transverse carbon magnetisation H_zC_x is generated and the chemical shift of carbon is evolved during t_1 while decoupling the deuterium spin using WALTZ decoupling and decoupling the proton using a simple π pulse. Another carbon $\frac{\pi}{2}$ pulse generates H_zC_z magnetisation which is allowed to relax for a time $T_{max} - T_1$ in order to remove H_z and C_z longitudinal relaxation components from the measurement. At point (e) this magnetisation is transferred to pure transverse hydrogen magnetisation and the signal is acquired from point (f) while decoupling the carbon spins using WALTZ decoupling.

For the phase cycling it should be noted that $\phi_{2/3}$ are cycled to remove unwanted coherences. Details of the delays and phase cycling used can be found in Section 5.3. Spectra are recorded at different T_1 values and the resulting spectra are fitted using FUDA,[160] written by D Flemming Hansen (last accessed at: http://www.biochem.ucl.ac.uk/hansen/fuda/), yielding an exponential relaxation rate R and the standard deviation of this rate based on the signal-to-noise of the spectra and the covariance from the fit of an exponential function to the derived intensities.

$^{13}CH_2D$ TROSY

The TROSY sequence used was taken from Ref. 82. This sequence is optimised to exploit the slow-relaxing carbon and proton transitions, i.e. $R^s_{2,C}$ and $R^s_{2,H}$. An HSQC sequence does not fulfil this purpose as it generates a mixture of coherences that are subject to fast relaxation. Specifically, the INEPT transfers of the HSQC are exciting the outer lines of the $^{13}CH_2$ triplet, as the central line does not evolve under scalar coupling. The standard HMQC, which is used in $^{13}CH_3$ methyl TROSY experiments, also does not perform well for $^{13}CH_2D$ groups due to the fast relaxation of the multiple quantum coherences during the indirect dimension evolution.[82]

Tugarinov and co-workers[82,87,88] have revised an optimised $^{13}CH_2D$ TROSY sequence that first uses a 'double' INEPT phase to transfer $\frac{1}{\sqrt{2}}$ of the proton magnetisation to the central line of the carbon-direct observed 1:2:1 triplet, where the magnetisation relaxes much more slowly ($R^s_{2,C}$) for big proteins. Subsequently different blocks are used to measure the relaxation of the following four coherences: D_+, D_z, $D_+D_z + D_zD_+$ and $3D_z^2 - 2$. These relaxation rates can be used to extract order parameters for the $C_{\gamma 1} - C_{\delta 1}$ axis,[74,82,163] which is discussed in more detail below.

Finally, magnetisation is transferred from the slow relaxing carbon central line to the slow relaxing proton transitions and both the sine and cosine modulated t_1 terms of magnetisation are kept and recorded in a similar manner as the sensitivity enhanced TROSY.[167,168]

The $^{13}CH_2D$ TROSY experiment generates combinations of in-phase (IP) and anti-phase (AP) terms that are recorded in two sets of data yielding one set with $IP^x + AP^y$ and another with $-IP^x + AP^y$ that have to be post-processed as detailed in[82] to yield a 'decoupled' CH spectrum of the methyl group. The data sets are added/subtracted, one is phase shifted by $\frac{\pi}{2}$ and the sets are again added/subtracted and finally circular shifted by $J_{CH} \approx 124\,\text{Hz}$ to yield the 'decoupled' absorptive spectrum. The processing script used for data post-processing can be found in the Appendix A.3.

Comparison of HSQC and TROSY Sequences

Albeit shorter acquisition time of the designed HQSC the signal-to-noise ratio of the resulting spectra is not always superior to the one of the TROSY sequence, which is a result of the fast relaxing transitions excited during the sequence and decoupling. I tested this by measuring the D_z relaxation rate $R^Q(D_z)$ with both sequences using the same overall acquisition time, i.e. same wall clock time to get the full set of data. Figure 2.15 shows a comparison of the resulting spectra. Comparison of Figure 2.15 (a) and (b) already suggests that

the TROSY sequence yields sharper peaks. Especially in areas where there are many peaks (see 'cross hair') the separation of these becomes difficult in the HSQC spectrum. Extraction of relaxation parameters from these peaks is then harder, as this relies on fitting the whole peak, preferably without overlap. Here Figure 2.15 (c) and (d) clearly show that the TROSY sequence is superior, i.e. it yields sharper peaks, especially in the carbon dimension.

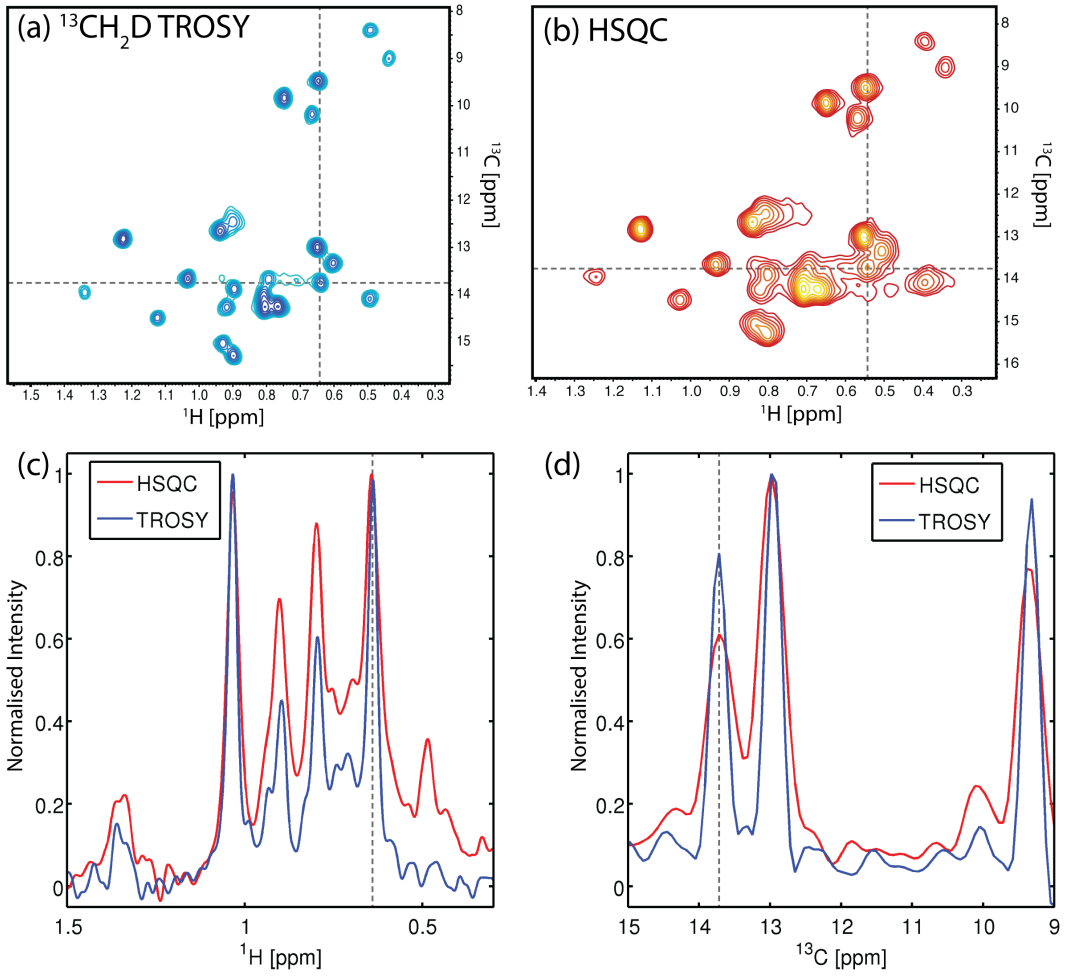


Figure 2.15.: Comparison of HSQC and TROSY spectra. (a) $^{13}CH_2D$ TROSY spectrum of HDAC8 wt. (b) HSQC spectrum of HDAC8 wt. Grey dashed lines represent the slices compared in (c-d). (c) Slice along the proton dimension. $^{13}CH_2D$ TROSY spectrum is coloured blue and the HSQC spectrum is coloured red. Grey dashed line marks the crossing with carbon dimension slice. (d) Carbon dimension slice. Colouring as in (c).

While the signal-to-noise ratio is not significantly improved using the $^{13}CH_2D$ TROSY over the HSQC, the improved resolution makes the TROSY the sequence of choice for relaxation measurements. Hence, I used the TROSY experiment to investigate the loop dynamics of HDAC8.

Deuterium Relaxation in $^{13}CH_2D$ Isotopomers

The relaxation of deuterium in methyl groups has been described in a series of publications.[71,74,85,163,169] Hence, only a brief overview is given here. To study side-chain dynamics of methyls using deuterium relaxation, the quadrupolar relaxation R^Q of different coherences has to be measured in order to map the spectral density function, yielding information about the order parameters of the methyl groups, i.e. yielding a measure for the side-chain motion on the ps to ns timescale.

In general, many sources of relaxation have to be taken into account. Following the description of Millet et al.[74] the Hamiltonian for relaxation H^R consists of a combination of the different contributions:

$$\begin{aligned} H^R =& H^Q(D) + H^D(CD) + H^D(I^iD) + H^D(I^jD) + H^D(I^iC) + H^D(I^jC) \\ &+ H^D(CC') + H^D(I^iI^j) + \sum_{k\neq i,j} H^D(I^iI^k) + \sum_{k\neq i,j} H^D(I^jI^k) \\ &+ \sum_{k\neq i,j} H^D(DI^k) \end{aligned} \tag{2.5}$$

H^D and H^Q are the contributions from the dipolar and quadrupolar interactions respectively. C, D, I^i and I^j denote the carbon, deuterium and proton spins of the methyl group, and C' is the adjacent carbon. I^k represents proximal protons, e.g. from the solvent or neighbouring amino acids. There are no neighbouring $^{13}C_{\gamma 1}$ carbons due to the used labelling schemes and the contri-

bution form external protons I^k will be very small. However, even in the case of only fractionally D labelled samples all terms except the quadrupolar term $H^Q(D)$ can be neglected to very good approximation,[74] i.e.:

$$R(H_zC_zD_z) - R(H_zC_z) \approx R^Q(D_z) \tag{2.6}$$

is fulfilled better than 2 %. The approximation 2.6 holds in the same way for the other measured coherences $D_+, 3D_z^2 - 2$ and $D_+D_z + D_zD_+$. For D_+^2 this error is much larger, as $R^Q(D_+^2)$ has no $J(0)$ contributions itself, but some dipolar process contribute.[74] Therefore, measuring $R^Q(D_+^2)$ requires semi-empirical correction of the acquired data and the error from the approximation $R(H_zC_zD_+^2) - R(H_zC_z) \approx R^Q(D_+^2)$ is $\sim 12\,\%$.[74]

Another error source for the measurement of the deuterium relaxation can arise in the case of chemical exchange. Here, deuterium relaxation measurements have the advantage that the contribution from chemical exchange is generally extremely small compared to the quadrupolar interaction and can safely be neglected, similar to the dipolar relaxation contributions. The chemical exchange contribution is proportional to $\Delta\omega$ up to $\Delta\omega^2$, where $\Delta\omega$ is the difference in chemical shift between the states.[170,171] Since $\frac{\gamma_H}{\gamma_D} \approx 6.5$ the influence on the relaxation from chemical exchange is much smaller than on proton relaxation rates. Moreover, the quadrupolar relaxation rates are much higher relative to proton relaxation rates, rendering the exchange contribution negligible.[82]

Hence, using the described pulse sequences that employ a $T_{max} - T_1$ element (Figure 2.14) the quadrupolar relaxation rates R^Q can be directly probes to extract order parameters as a measure of the dynamics of the side-chains.

Extracting Order Parameters from Deuterium Relaxation

To extract Lipari Szabo order parameters[89,90] from the measured relaxation rates the spectral density function has to be mapped and subsequently modelled

using an adequate model. The quadrupolar relaxation rates R^Q measured depend on the spectral density $J(\omega)$ as follows:[74,172]

$$R^Q(D_z) = \frac{3}{40}\left(\frac{e^2qQ}{\hbar}\right)^2 [J(\omega_D) + 4J(2\omega_D)] \quad (2.7)$$

$$R^Q(3D_z^2 - 2) = \frac{3}{40}\left(\frac{e^2qQ}{\hbar}\right)^2 [3J(\omega_D)] \quad (2.8)$$

$$R^Q(D_+) = \frac{1}{80}\left(\frac{e^2qQ}{\hbar}\right)^2 [9J(0) + 15J(\omega_D) + 6J(2\omega_D)] \quad (2.9)$$

$$R^Q(D_+D_z + D_zD_+) = \frac{1}{80}\left(\frac{e^2qQ}{\hbar}\right)^2 [9J(0) + 3J(\omega_D) + 6J(2\omega_D)] \quad (2.10)$$

Where $\frac{e^2qQ}{\hbar}$ is the quadrupolar coupling constant, which is 167 kHz in a CH_2D group.[74,173] Formulas 2.7 to 2.10 show that measuring the four relaxation rates $D_z, 3D_z^2 - 2, D_+$ and $D_+D_z + D_zD_+$ allows one to map three points of the spectral density function at frequencies $0, \omega_D$ and $2\omega_D$, where ω_D is the deuterium Larmor frequency. I then used the model free approach from Lipari and Szabo,[89,90] as described in Section 1.3.3. Specifically, the simple 2 parameter model (LS2) that assumes that the $C_{\gamma 1} - C_{\delta 1}$ bond vector motion is faster than the motional tumbling of the protein. This spectral density model can be written as:

$$J(\omega) = \alpha S_f^2 \frac{\tau_R}{1 + (\omega\tau_R)^2} + \left(1 - \alpha S_f^2\right) \frac{\tau}{1 + (\omega\tau)^2} \quad (2.11)$$

Here, S_f^2 is the order parameter separating the isotropic overall tumbling τ_R and $\tau^{-1} = \tau_R^{-1} + \tau_f^{-1}$, where τ_f is the correlation time of the fast internal motion of the $C_{\gamma 1} - C_{\delta 1}$ bond, which is dominated by the fast methyl rotation. Hence, S_f^2 describes the amplitude of internal motion. α is a geometric factor that takes into account the geometric relation of the $C_{\delta 1} - D$ bond to the 3 fold symmetry axis of the methyl rotation, i.e. $\alpha = (3\cos^2\Theta - 1)^2/4 = 0.1108$ with $\Theta = 109.5°$ the angle between the 3 fold symmetry axis of the methyl and the $C_{\delta 1} - D$ bond.[74]

The LS2 model (Equation 2.11) can be used to globally fit the relaxation rates

(Equations 2.7 to 2.10), which I obtained using MATLAB (Appendix A.5). In brief, a global fit is achieved by calculating χ^2 values for relaxation rates, which are then concatenated to a single vector and minimised. To test the robustness of and estimate the error of the solution of the fit is repeated with normal-distributed values for the relaxation rates, i.e. for each repeat (typically $n_{rep} = 100$) the relaxation rates are taken randomly from the normal distribution with the expectation value $\mu = R_{best\,fit}$ and σ as the standard deviation of the relaxation rate fit. Essentially, performing a Monte-Carlo error estimate. Since the system is overdetermined it allows to fit S_f^2 and τ, i.e.the fast motion τ_f as long as τ_R is known from other experiments such as amide relaxation experiments. Here I chose $\tau_R = 35\,\text{ns}$ such that the order parameters are below 1, which is largely in line (albeit higher) than estimates from $^1\text{H} - ^{15}\text{N}$ experiments. All other parameters can be found in the Appendix A.5.

2.4.3. Effect of Lys33Glu Mutation

Using $^{13}CH_2D$ TROSY as described I aimed to investigate the L1:L2 interaction and possible changes in conformation and dynamics of these loops by the introduction of a single point mutation as described in Section 2.1. Figure 2.16 shows a zoom in to the $^1\text{H} - ^{13}\text{C}$ spectrum of the Ile side chains, focusing on the peaks that shift upon the Lys33Glu mutation. Indeed, most of the side-chain peaks of the Lys33Glu mutant overlap with the ones of the wild type. The most interesting peak is the one of Ile34 (Figure 2.16) as it is located right next to Lys33. Unfortunately it is unobservable in the mutant spectrum. This is most likely due to the lower signal-to-noise ratio of the HDAC8 Lys33Glu spectrum relative to the HDAC8 wt one, which is a result from a lower sample concentration, it is possible that the Ile34 peak is now overlapping with some of the other peaks or broadened due to chemical exchange on the $\mu\text{s} - \text{ms}$ timescale. There are two distinct shifts observable: Ile94 and Ile284. The resonance of Ile284 exhibits a small shift in the carbon dimension. Ile284 is located in

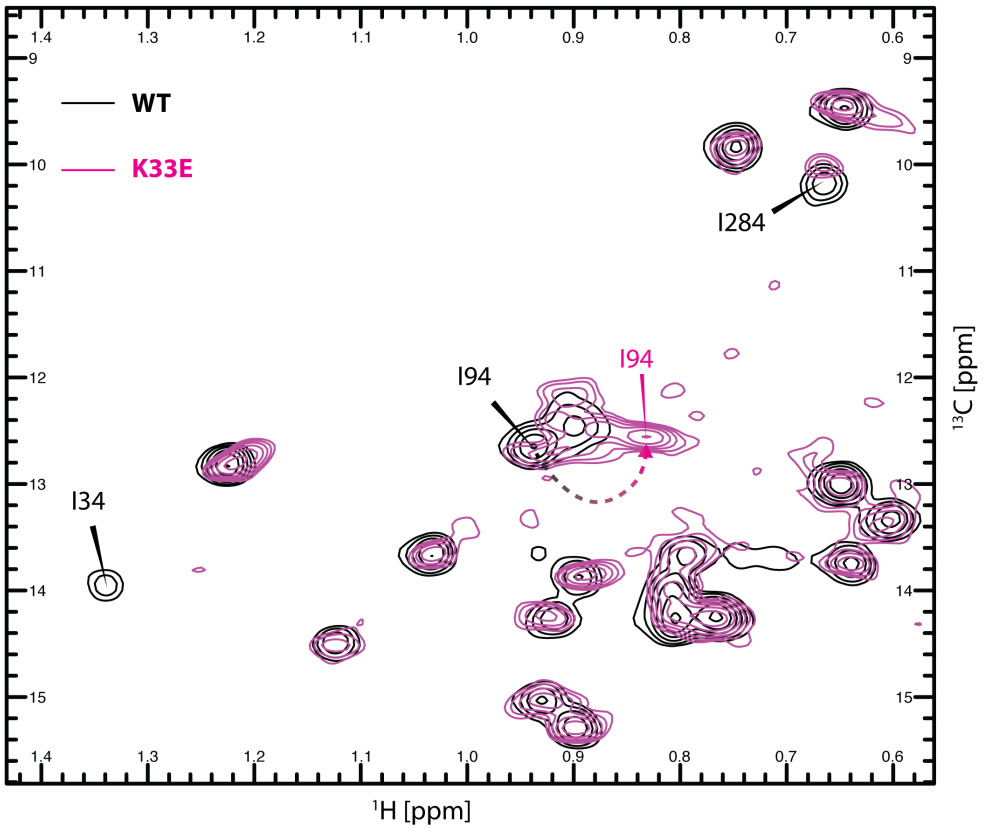


Figure 2.16.: $^{13}CH_2D$ TROSY of Ile $C_{\delta 1}$ methyl labelled HDAC8 wt (black) and Lys33Glu (magenta) mutant. Arrows highlight the peaks and peak shifts discussed in the text. Buffer conditions and parameters for the experiments can be found in Section 5.3.

the hydrophobic sandwich of helices and beta sheets (Figure 2.12); however since other residues in this area do not exhibit shifts, the structural change seems small, but might still cause interference with the enzymatic function. The Ile94 resonance seems to undergo a distinct shift. The assignment in the mutant spectrum has not been confirmed by single point mutations, but relative signal intensity, line shapes and distinct relaxation characteristics support the assignment. The shift of the Ile94 is interesting as the $C_\alpha - C_\alpha$ distance of Lys33 and Ile94 is larger than 2 nm, suggesting that the loop L1:L2 interaction

in HDAC8 wt is indeed present and allosterically modulated by the mutation of Lys33.

2.4.4. Side-Chain Dynamics of Ile Residues

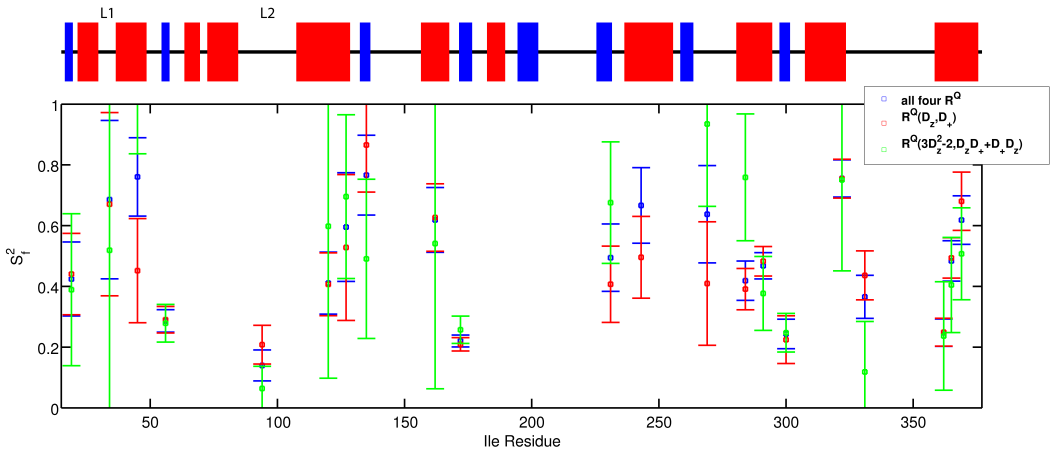


Figure 2.17.: S_f^2 LS order parameters for the $C_{\delta 1}$ methyl group of observable Ile residue of HDAC8 wt based on deuterium relaxation measurements. The graphic on the top represents the secondary structure of HDAC8: black lines indicate no secondary structure, blue boxes indicate position of beta sheets and the red boxes represent α helices. L1 and L2 loop positions are highlighted. The graph shows the S_f^2 including estimated error as achieved by globally fitting all four measured relaxation rates R^Q of the coherences $D_z, D_+, 3D_z - 2$ and $D_+D_z + D_zD_+$ in blue. Fits of S_f^2 based only on R^Q of D_z and D_+ are shown in red and fits based on relaxation rates of the $3D_z - 2$ and $D_+D_z + D_zD_+$ coherences are shown in green.

Next to the structural information is the dynamical information, i.e. the flexibility of the loops in the wt and mutant enzyme. Using deuterium relaxation data I calculated the Lipari Szabo order parameters S_f^2 for all observable Ile resonances. Figure 2.17 shows S_f^2 order parameters calculated from the relaxation rates of the four measured coherences. This shows that Ile94, which is located in the L2 loop, with $S_f^2 \approx 0.15$ is very dynamic, as is expected for a residue in a dynamic loop. Ile34 seems less dynamic with a higher S_f^2, however,

the error for this value are large, which is a direct result of the overall weak signal of this peak. While the peak shape of Ile34 (Figure 2.16) suggests that this side chain is dynamic on a μs $-$ ms timescale it seems not to be very dynamic on the here probed timescale, ps to ns. The data quality hinders unambiguous characterisation of the L1 and L2 loop dynamics and overall can only serves a rough estimate of the side-chain dynamics for most residues.

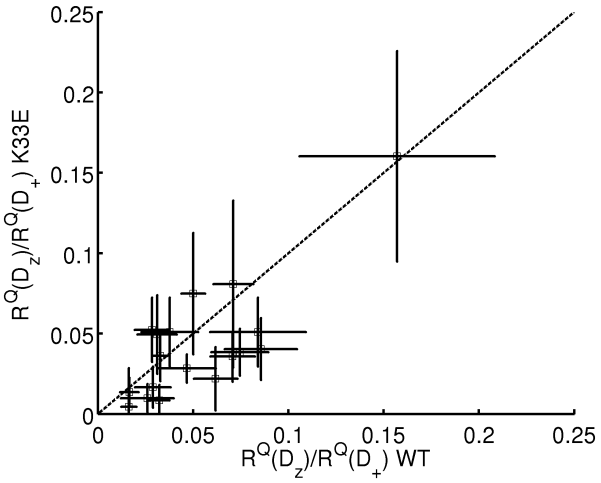


Figure 2.18.: Deuterium relaxation rate ratios $R^Q(D_z)/R^Q(D_+)$ of the deuterium in the $C_{\delta 1}$ methyl group or each observable Ile residue. Wild type ratios are shown on the x-axis and the Lys33Glu ratios are shown on the y-axis, both with their associated error.

To detect possible changes of the dynamical content of the protein, specifically the L1 and L2 loop, the deuterium relaxation rates in the Lys33Glu mutant of HDAC8 was measured since this is the most deleterious mutant. As mentioned in Section 2.4.3 the Lys33Glu mutants lower expression level results in a lower signal-to-noise ratio relative to the wt, which already has large errors associated with the measure relaxation rates. In fact, it was not feasible to measure the relaxation rates of the $3D_z - 2$ and $D_+D_z + D_zD_+$ coherences due to the poor signal. Figure 2.18 shows a comparison of the $\frac{R^Q(D_z)}{R^Q(D_+}$ ratios of the Lys33Glu mutant and the wild type enzyme. Again, the associated errors

are large and changes of the relaxation rate ratios are hard to unambiguously interpret. While the data shows small trends, an interpretation of these is speculative at best.

2.4.5. Conclusion of NMR Results

The NMR data presented suggests that the L1 and L2 loop are dynamic, at least the side chains of the observable isoleucine amino acids in these loops. The peak of Ile34 observed in carbon proton spectra is broadened, hinting at a dynamic character of this side chain. However, deuterium relaxation data suggests that this side chain is not very dynamic on the accessible ps to ns time scale. Hence, the side is most likely exhibiting motions in the μs to ms time scale or higher. The Ile94 peak is very sharp yielding good quality data. The low S_f^2 order parameter shows that this side chain is highly dynamic. Dynamics on the μs to ms time scale can be measured using residual dipolar coupling experiments, which could give great insight into the dynamics of the L1 loop.

Moreover, the comparison of CH_2D TROSY spectra of wt and mutant shows that the Lys33Glu mutation leads to specific changes, most notably of the Ile94 peak. This strongly suggests that there is an L1:L2 interaction present in the enzyme and more importantly that there is communication between the L1 and the L2 loop, which has not been grasped by crystal structures or other studies. This also supports the working hypothesis I derived from MD simulations (Section 2.1). The peak of Ile284, however, also exhibits a small shift, which should not be neglected as it could hint at other changes of the structure that might explain or at least add to the explanation of how the mutation effects enzymatic activity.

The study shows that investigating the dynamics of HDAC8 isoleucine is possible. But to decipher the influence of the loop dynamics and interaction on the enzymatic activity of HDAC8 better quality data is needed.

2.5. Probing a Transient Product Release Tunnel using MD

In the past and recent HDAC8 literature it has been hypothesised that there exists an exit channel for the acetate product.[19,103,126,174,175] While there are different proposed tunnels[19,103] they all furcate at the internal cavity under the L1 loop, which has been shown to be occupied by inhibitors such as TSA (PDB: 1T64), i.e. form a structure that is better described as a pocket. This idea has led to designing specific HDAC8 inhibitors that specifically target this internal cavity or pocket. Targeting this pocket seems to be successful for improving the specificity[176] of the drugs, which is most likely a direct result of the differences of these pockets amongst the class I HDACs.

However, no direct observation or proof exists to whether this pocket in fact acts as an product release tunnel. This cannot be as easily seen from static crystal structures and hence is only a working hypothesis. Here, I will present supporting evidence for the acetate release tunnel hypothesis including an identification and characterisation of the exit channel for HDAC8. I analysed MD data for dynamic channel openings and the transport of water along this tunnel. In principle, a single acetate molecule at the catalytic site should be used to explore the transport of this product. However, in unbiased simulations this single transport event would most likely take too long to occur on an MD time scale accessible using current high-end resources. While steered MD could be considered, without knowledge of the position of the tunnel achieving converged sampling is difficult. In contrast, in the simulation presented in Section 2.1, approximately 17.000 water molecules are present that should be able to permeate any transient tunnels.

Thus, no acetate is used in the simulations presented. However, in conjunction with other studies, especially a recent study of the acetate exit channel in HDAC-1 and -2,[175] the characterisation of the exit tunnel here yields a very viable explanation for acetate transport. Wambua et al. could show that

mutating residues that were thought to be involved in the acetate release tunnel in HDAC-1 and -2 change the activity and product inhibition by acetate.[175] Hence, the presented observations yield a viable mechanism for the acetate release hypothesis.

2.5.1. Analysis of MD Trajectory

Apart from the substrate binding channel, crystal structures yield no direct observation of channel-like connections of the internal cavity to the bulk solvent. Hence, if there were such a tunnel it would be transient in nature, i.e. it has to form dynamically. Not only might such dynamic tunnel formation be visible without acetate, but other researchers have speculated that such a tunnel in HDAC-1 and -2 could have an effect on the enzymatic activity of HDACs by transporting water molecules to or away from the active site.[175]

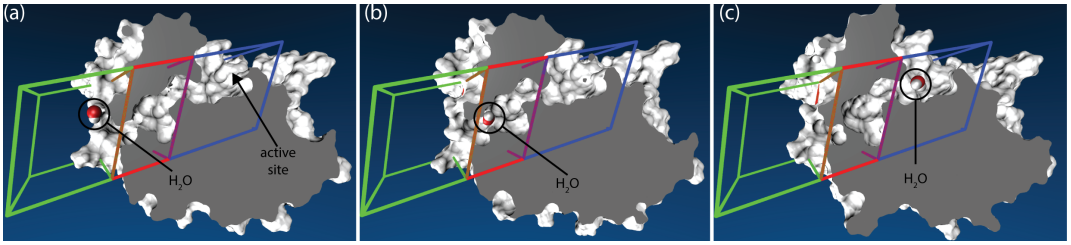


Figure 2.19.: Frames of an 'in' transition (from bulk into active site), illustrating the transient tunnel of HDAC8. Water molecule in transit shown as red and white beads. HDAC8 is shown as a grey surface representation 'sliced' by a clipping plane that lies within the transient tunnel in the presented frame. The green (box 3), red (box 2) and blue (box 1) boxes are visual representation of the boxes used in the water transport algorithm. Water molecule position and active site position are highlighted. (a) Water molecule is in the bulk solvent. (b) Water molecule enters the channel via the His42:Tyr18 gating. (c) Water molecule at the end of the transition near the active site. The elapses MD time of the shown transition is 800 ps.

Following this argument I used a combination of programmes and analysis routines to study the formation of transient tunnels and water transport in these

tunnels between active site and bulk. At first I used the HOLE programme[177] on frames taken at regular intervals (50 ps for the final analysis) of the trajectory, which yielded the first frames where a tunnel had formed that was capable of encompassing a water molecule at the narrowest point ($r_{min} > 1.15$ Å).[177] HOLE measures the inner dimension of pores in protein. The programme will try to find a pore along a given general direction by measuring the distance to van der Waals contacts along this direction. This is done in increments, where the centre of the measured pore is adjusted every step to align with the point that is in the pore and furthest away from all protein atoms. However, since HOLE is intended to be used to study pore proteins it has some inherent limitations for studying the narrow and transient tunnel of HDAC8. It turned out that the tunnel in HDAC8 can be very narrow or closed at different sites since part of it is an internal cavity, which poses problems for the HOLE analysis when searching for the smallest possible radius of the tunnel at any given time. HOLE is also not designed to track water molecules along a pore or tunnel over a MD trajectory. Tracking possible water transport is not entirely trivial in a simulation containing $\approx 17,000$ water molecules. To achieve this I wrote MATLAB routines that first assign each water molecule position to boxes that best describe the transient tunnel and are located at the active site (blue / box 1), between internal cavity and bulk (red / box 2) and the bulk exit/entry point (green box 3) as depicted in Figure 2.19. So as not to move the defined boxes for each frame the structures (protein and solvent water) are aligned to the protein backbone beforehand. The box assignment is then achieved by simply defining the three boxes in Cartesian coordinates and checking whether each water molecule within each frame is in one of the boxes or outside and subsequently 'labelling' it with the box number or a zero for each frame. This results in a 'box trajectory' for each water molecule that either has a value of 0, 1, 2 or 3 for each frame (or time point).

Subsequently these 'box trajectories' are analysed for 'in' or 'out' transitions,

i.e. whether a trajectory contains a $1 \rightarrow 2 \rightarrow 3$ pattern ('out' transition) or a $3 \rightarrow 2 \rightarrow 1$ transition ('in' transition). One has to take into account that a molecule can be inside a box for several frames and that it may transition between box 1 and 2 for several times before it decides to make a final 'out' transition to 3, e.g. $1 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 3$. This is achieved by a combination of 'if' and 'while' statements as can be seen in Appendix A.4.

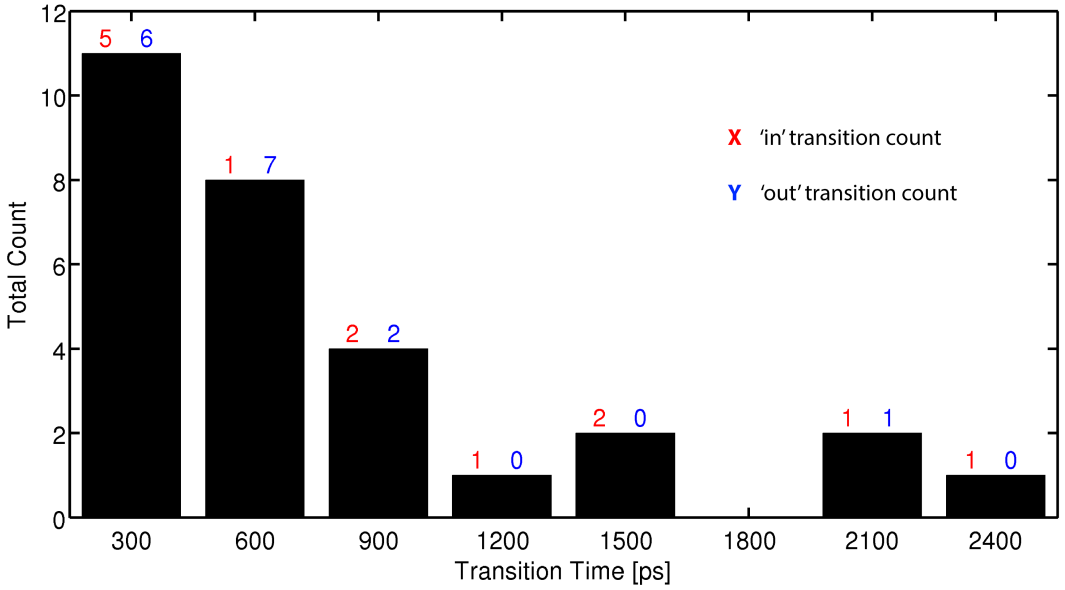


Figure 2.20.: Counts of 'in' an 'out' transition times during the simulation. Black bars mark the total counts ($n_{in} + n_{out}$) over the transition time, i.e. the time it takes for a water molecule to go from box 1 to box 3 ('out') or from box 3 to box 1 ('in'). The numbers above the bars represent the number of 'out' (n_{out}) and 'in' (n_{in}) transitions

This step is followed by a short overlap check, i.e. the transitions are analysed against each other so that trajectories such as $1 \rightarrow 2 \rightarrow 3 \rightarrow 2 \rightarrow 1$ are not counted as two transitions, one 'in' one out. These cases are usually not real transitions as there can be some cases where the alignment of the protein is unfavourable relative to the box positions, putting the water in box 3 very briefly even though it is still not right at the active site. The programme

produces TCL scripts for VMD to visually inspect every transition. Figure 2.20 shows the number of transitions found during the simulation analysed.

2.5.2. Tunnel Characterisation

The pocket or internal cavity under the L1 loop can be occupied by inhibitor molecules in crystal structures (e.g. PDB 1T64) and other studies suggest that acetate is capable to interact with this site.[19,174] From previous studies it was not clear how and where exactly the acetate would be released. Using the transient tunnel analysis as described above, 29 transitions (n_{in} = 13 and n_{out} = 16) of water molecules were observed showing which residues are gating the cavity and the bulk water (Figure 2.20). An example of such a transition is shown in Figure 2.19 and Video 3 (Appendix A.11). The water transition is on the 100 ps time scale and since there is a similar number of 'out' and 'in' transitions observed, it seems to be a mere diffusion process and not directed transport.

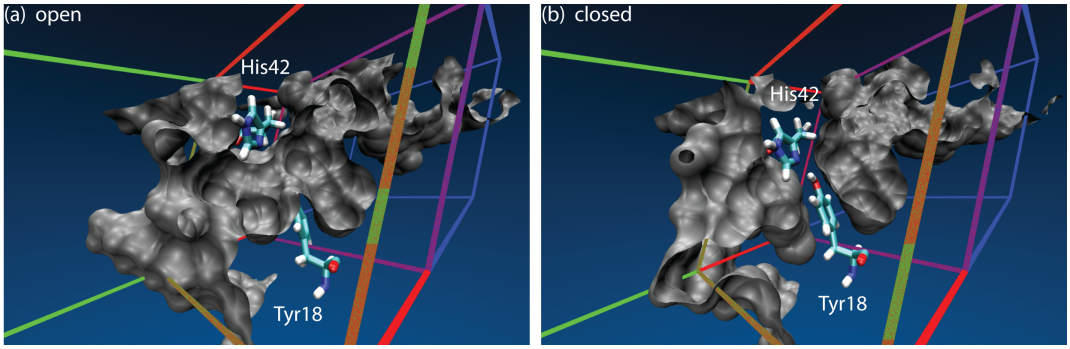


Figure 2.21.: Illustration of the tunnel gating by Tyr18 and His42. Transient tunnel is represented with a grey surface represent and Tyr18 and His42 are highlighted in a stick representation. (a) Closed tunnel with H_η(Tyr18) and N_δ(His42) interaction. (b) Open tunnel with no interaction of Tyr18 and His42 present.

All transitions between the bulk and the internal cavity happen at the same location and seem to be gated by His42 and Tyr18 (Figure 2.21), which is in

agreement with the hypothesis from Vannini et al.[103] For clarity the gating mechanism derived from the MD simulations is depicted in Figure 2.22.

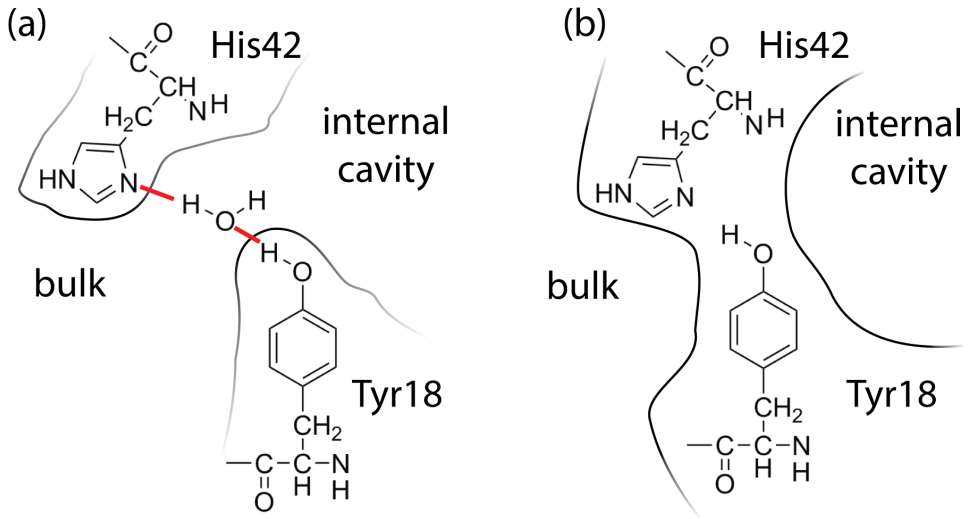


Figure 2.22.: Illustration of the gating of water molecules during transit. Gating residues are represented with their chemical structure. Black lines illustrate the solvent accessible surface area. (a) Open channel with a water molecule between the gating residues, forming hydrogen bonds (red) to the N_δ(His42) and H_η(Tyr18). (b) Closed channel where H_η(Tyr18) and N_δ(His42) form a hydrogen bond.

However, this gating site is not always the most narrow one of the overall tunnel; Arg37 can also gate the internal cavity from the substrate tunnel as proposed by Haider et al.[19] Furthermore, a recent study investigating the acetate tunnel hypothesis using single point mutations and enzymatic assays of HDAC1, a class I HDAC, identified the Tyr23 and Tyr24 motif to gate the internal cavity:bulk interface in HDAC1.[175] In their analysis using structural alignments they do not pick up the site I identified as the gating site, i.e. Tyr18 and His42 in HDAC8, even thought both gating sites are at a similar position. There are, however, conserved residues that are proposed to act as hydrogen bond partners for the acetate when transported in the internal cavity, Arg37 and Cys153 in HDAC8 are strictly conserved in class I and II HDACs,[103] while mutating those residues in HDAC8 or HDAC1 have been shown to reduce en-

zymatic activity by 4 to 50-fold and the product inhibition of acetate by 2 to 3-fold. In fact, Cys153Phe is a missense mutation that has been proposed to be related to acetate release, i.e. blocking a release channel, in a very recent study.[124]

2.5.3. Conclusion and Discussion

A transient water tunnel in HDAC8 has been identified in MD simulations, which agrees well with previous experimental observations and hypotheses regarding acetate release. While only water molecules are observed, it seems very likely that a small molecule such as acetate would be able to pass through this tunnel by e.g. inducing larger movement of the gating amino acids Tyr18 and His42. In comparison to the study concerning HDAC1[175] it becomes clear that a) the mechanism of transport seems to be very similar amongst class I HDACs and b) that there are significant differences in the structure of the internal cavity of these isoforms. This again highlights the potential for structure based drug design to target this cavity as the different shapes and sizes of these internal cavities are most likely to yield isoform specific inhibition.[176]

Moreover, the tunnel identified here accounts for the phenotype of the missense mutation Ile19Ser in HDAC8 involved in CdLS.[121,123] This mutation is directly C-terminal to the gating residue Tyr18, hence the 5-fold decrease in enzymatic activity[123] is likely to be an effect on the gating mechanism, while 5-fold decrease in enzymatic activity of this mutation[123] is similar to what is observed for Tyr23 and Tyr24 mutations in HDAC1.[175] Indeed, His42Pro is also a CdLS associated missense mutation.[121,123] However, HDAC8 His42Pro expression trials are reported not to yield any soluble protein,[123] which is perhaps to be anticipated, since His42 is located in an α-helix where a mutation to Pro is expected to yield significant disruption of the protein structure.

All in all, the observations fit in well with published experimental data and the working hypothesis in the field. Therefore, the proposed mechanism seems

very viable. While there might be other possible transient channels, only one type of mechanism is observed, however this does not exclude the possible presence of other tunnels that have not been observed due to sampling issues etc. However, with almost 30 observed transitions the mechanism observed seems to be robust.

There are several possibilities to test the presented mechanism further. Computationally the transport of acetate could be studied using steered MD methods to gauge the free energy landscape of acetate transport along this tunnel. From an experimental point of view the activity of single points mutations of tunnel residues could be studied. Especially the gating residue Tyr18 would be a good candidate. Mutation of this residue to e.g. Phe would remove the OH group responsible for the hydrogen bonds to the water molecule during gating (Figure 2.22 (a)). A Tyr18Phe mutations would also be relatively conservative, i.e. it is unlikely to dramatically change the protein stability due to the similarity of Tyr and Phe.

3. Collaborative Projects

In the following I will describe published work I contributed to during my time as a PhD student, which does not constitute my main project concerning the HDAC8 loops. I will highlight my part of the work in detail in the context of the project it was part of. In that respect I am very grateful for all the great scientists I could work with during these projects.

3.1. Dynamics of the Metal-Site of Paramagnetic Metalloproteins from NMR Parameters Using Natural Bond Orbitals

The functionally important part of metalloproteins often is the metal-site, where the detailed geometric and electronic structure are key for the proteins function. Metal-sites in proteins can be studied by crystallography, X-ray absorption spectroscopy[178,179] or EPR,[180] yielding information about the geometric and electronic structure in a crystal or frozen glass conditions.

Especially for flexible proteins, however, it is desirable to investigate the dynamics and structure of the protein (especially of the metal-site) simultaneously in aqueous solution.

A study where paramagnetic NMR was used in conjunction with MD simulations to study the structure of the iron-sulfur protein rubredoxin (from *C. pasteurianum*) is presented. While in the study the blue copper protein plastocyanin was studied alongside rubredoxin, this Section will focus only on rubredoxin as a) only here I contributed by using MD and restrained MD

to investigate the metal-site structure of the protein and b) plastocyonin was studied in direct analogy, where the details and results can be found in Ref. 2.

Paramagnetic NMR yields a wealth of information about the electronic and geometric structure of a metal protein,[181–183] but is usually limited by an accurate description of the unpaired electron density. The electron spin density is usually spread over metal coordinating residues.

Hence, there is a strong anisotropy that has to be accounted for, which lets the typical point-dipole approximation fail. In the point-dipole approximation the electron spin is approximated to be only located in a point at the metal position. To alleviate this a method was developed to use some fast DFT calculations and use natural bond orbitals (NBOs) to accurately describe the electron spin density at the metal-site for later comparison and interpretation of experimental data.

In fact, the description using NBOs is accurate enough to be effectively used to restrain MD simulations, driving the simulated structure towards a native structure.

3.1.1. Paramagnetic Relaxation Enhancement

The primary source of information can be extracted by measuring the paramagnetic relaxation enhancement (PRE) of certain nuclei. The PRE contains contributions stemming from the Fermi coupling, the Curie spin relaxation and the dipolar relaxation. The relaxation enhancement from the Fermi shift can be neglected as it is approximately 0.5 % of the dipolar coupling for all nuclei considered here.

Hence the PRE for longitudinal relaxation can be written as:[2,184]

$$R_{1p} = \frac{6}{5}\left(\frac{\mu_0}{4\pi}\right)^2 g_e^2 \mu_B^2 \gamma_I^2 r_{eff}^{-6} \left[\langle S_z\rangle^2 \frac{\tau_R}{1+\omega_I^2\tau_R^2} + \left(S\left(S+1\right)/3 - \langle S_z\rangle^2\right)\frac{\tau_{c,1}}{1+\omega_I^2\tau_{c,1}^2}\right] \tag{3.1}$$

$$= \xi r_{eff}^{-6} \tag{3.2}$$

where g_e is the isotropic electron g-factor, S is the electron spin number, μ_B is the Bohr magneton, γ_I is the gyromagnetic ratio of the studied nucleus, μ_0 is the magnetic permeability of free space, $\tau_{c,1}$ is the correlation time of the electron-nucleus dipolar interaction vector, τ_R is the rotational correlation time of the protein and $\langle S_Z\rangle$ is the thermal average of the electron spin magnetisation.[184]

The Curie spin relaxation is encapsulated in Equation 3.1 as the τ_R dependent component. Often this contribution can in fact be neglected.[184] In the presented case for the protein rubredoxin the contribution to R_{1p} from Curie relaxation is about 2.5%.[2]

Cross-relaxation terms from Curie and dipolar relaxation can also be neglected as the magnetisation of the measured nuclei (H and N) relaxes faster than the scalar coupling between the atoms and the relaxation of nitrogen is measured in-phase in respect to H.

While neglecting Curie relaxation simplifies Equation 3.1 further, the effective distance of the nucleus and electron density r_{eff} has to be taken into account, i.e. the distribution of electron spin density $\rho(\mathbf{r}) = \rho_\alpha(\mathbf{r}) - \rho_\beta(\mathbf{r})$, which is distributed on the metal and ligand atoms. The effective nucleus-electron distance is given by:

$$r_{eff}^{-6} = \frac{4\pi}{5}\sum_{\nu=-2}^{2}\left|\int_V \mathrm{d}\mathbf{r}\hat{\mathcal{F}}_2^\nu(\mathbf{r}-\mathbf{r}')(\rho_\alpha(\mathbf{r}) - \rho_\beta(\mathbf{r}))\right|^2 \tag{3.3}$$

where $\hat{\mathcal{F}}_2^\nu = ||r||^{-3}Y_2^\nu(r/||r||)$ are the spatial components of the nucleus-electron

dipolar operator centred at the nucleus at position $\mathbf{r}'$. $Y_2^\nu(r/||r||), \nu = -2, .., 2$ are the spherical harmonics. Essentially Equation 3.3 describes an integration over the interaction of the nuclear spin with the whole delocalised electron spin density.

Hence, in the effective distance a lot of information is encapsulated about the geometrical structure, by the means of nuclear position, and electronic structure, by means of unpaired electron spin density. From a structural perspective it is clear that one is interested in the positional information that can be obtained from the measurement of nuclear relaxation rates.

3.1.2. Hyperfine Shift

The unpaired electron spin density at or near a nucleus does not only cause relaxation enhancement (at least if not in an s-orbital), but also causes a resonance shift. This is mainly caused by the unpaired electron spin density at the nucleus, a Fermi contact shift, or by unpaired electron spin density in the direct vicinity, the pseudo contact shift.

While only a fraction of the unpaired electron spin density resides at he nucleus of interest, even small densities can have a significant effect on the relaxation and chemical shift due the spatial closeness of the spins. For nuclei close to the metal-site the Fermi contact shift is usually much larger than the dipolar pseudo contact shift. Using the crystal structure of the metal-site of oxidised rubredoxin (PDB id 5RXN) the upper limit for the pseudo contact shift can be calculated from the magnetic susceptibility of Fe(III) rubredoxin.[185] This shows, that the pseudo contact shift is at maximum 8.5% of the overall hyperfine shift.[2] Hence, the measured shift is dominated by the Fermi contact shift, which is directly proportional to the unpaired electron spin density at the nucleus position.[2]

Following this, it is clear that measuring of the hyperfine shift δ_{hf}, which is es-

sentially the peak shift for diamagnetic versus paramagnetic Fe(III) rubredoxin, allows a direct observation of the unpaired electron spin density at the nucleus position. The diamagnetic peak position here was estimated from a random coil value. Since the contribution of electron spin density in the local orbital of the relaxing nucleus does not bear information of the nucleus-electron positioning, the aim was to measure the contact shift and later correct the measured PRE to only report on the positional components.

DFT calculations for this study have been done at the UB3LYP functional and the 6-31G* basis set, using Gaussian03 (Schrödinger LLC).

3.1.3. Natural Bond Orbitals

A description of the unpaired spin density in a metal protein in terms of DFT is rather complicated and unintuitive due to the often large number of basis functions used to describe the electronic structure.

Natural bond orbital (NBO) analysis can be seen as a means of transforming DFT of e.g. Gaussian type atomic orbitals into a natural bond orbital picture. The one-particle density matrix of two atoms is transferred into the NBO basis set and only natural bond and core orbitals are considered, where the eigenvector of the transformation including the occupation number (over a threshold) constitutes what is called a pre-NBO (pNBO). The NBOs are then the orthogonalisation of the pNBOs. The mainly occupied pNBOs for the small model system of rubredoxin are illustrated in Figure 3.1.

3.1.4. NBO and Electron Spin Density

Using NBO analysis on a model system of rubredoxin (Figure 3.1) it turns out that the unpaired electron density can be described 90% accurately only using a few pNBOs.[2] This can be seen in Figure 3.2 (a) and (b), where the number

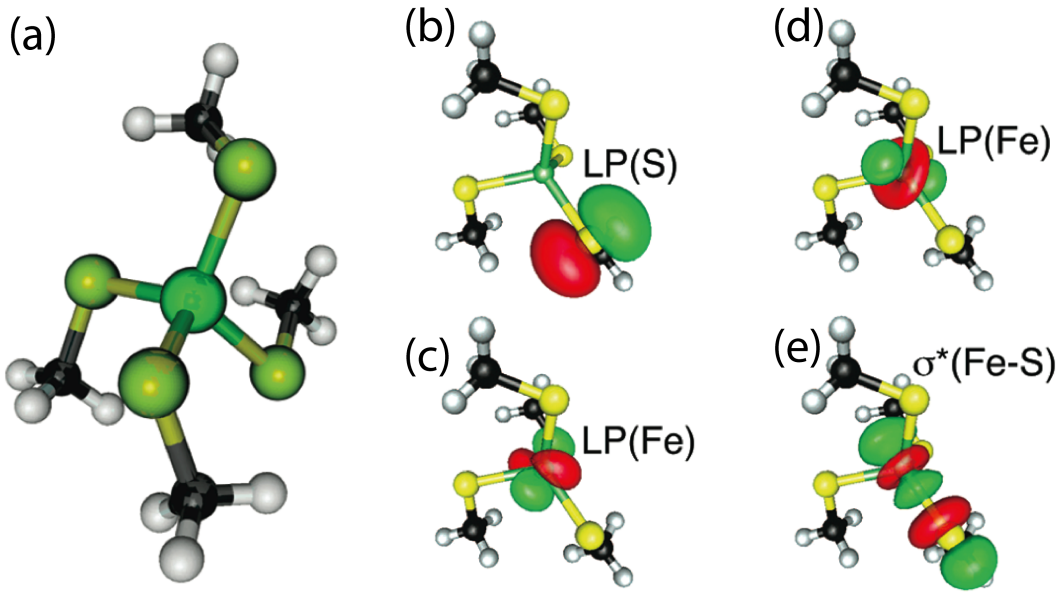


Figure 3.1.: Model system used for the NBO analysis of rubredoxin and mainly electron spin occupied orbitals. Iron atom is shown as a green bead, sulphur atoms are shown as yellow beads, carbon atoms are shown as black beads and hydrogens as white beads. Hydrogens are added to cap the valency of the model system. Electron spin density occupation is shown by green/red orbitals. (a) Picture of the model system; (b) Lone pair at the sulphur; (c) Lone pair at the high spin iron; (d) Another lone pair at the iron atom; (e) σ^*(Fe-S) antibond. Occupation numbers of the orbitals can be seen in Figure 3.2. Panels reproduced from Ref. 2.

of pNBOs used to describe the unpaired electron density is small compared to the number of Gaussian type orbitals description.

Hence, it seems like it is possible to use a rather small system for DFT calculations and subsequently only a few pNBOs to describe the unpaired electron spin density rather well, i.e. mainly the Fe and S lone pairs (LPs) and the σ^*(Fe-S) anti bond orbitals. During this project it could be shown that the PRE of hydrogens can effectively be calculated from the electron density described by the few pNBO and Equation 3.3.[2]

This is only possible as local unpaired electron density at hydrogens occu-

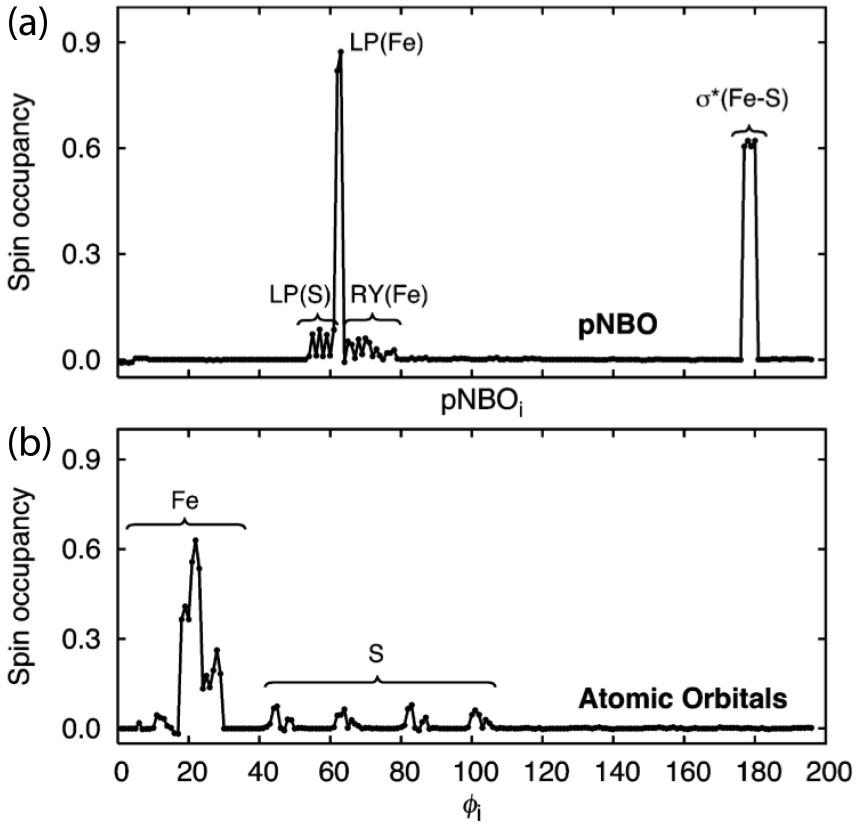


Figure 3.2.: Comparison of orbitals occupied with unpaired electron spin density. (a) The pNBO occupation and (b) the occupation of Gaussian type atomic orbitals. Clearly in the pNBO description there are fewer orbitals occupied, where only a few (LP(S), LP(Fe) and σ^*(Fe-S)) are dominating compared to a much broader distribution for Gaussian type orbitals. Figure reproduced from Ref. 2.

pies s-orbitals, which only contribute to contact shift but not to relaxation enhancement. Hence, the contact shift does not have to be considered here and distance information of the hydrogens and the metal-site can be extracted straightforward from a relaxation measurement.

Yet, these measurements have several pitfalls due to the amount of broadening of paramagnetic proton NMR. Here, the relaxation enhancement often broadens the proton signal beyond detection. Therefore, we investigated whether it would be possible to use the simple NBO picture from above for e.g. amide nitrogens in rubredoxin.

3.1.5. PRE of Nitrogen Atoms

For ^{15}N the calculation of the PRE from the unpaired electron spin is more complicated as nitrogens have p-orbitals where unpaired electron occupation would have a significant impact on the PRE due to the closeness, even though the occupation is only around 0.03%.[2]

This can be seen by a comparison of the calculated contact shifts of DFT results for a much larger system (104 atoms in the vicinity of the metal-cite). In essence, for nitrogen atoms that are not part of the small model system, local NBOs (LNBOs) can be found to describe the local unpaired electron density at the nucleus position. Which LNBOs have to be used is defined by a spin polarisation pathway.[2]

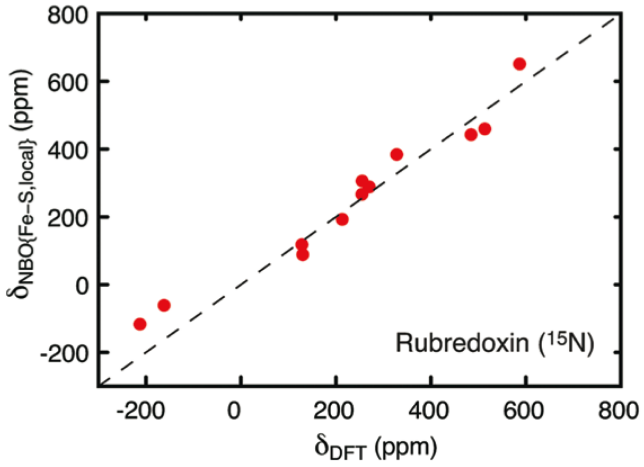


Figure 3.3.: Contact shift calculated from pNBOs and LNBOs are plotted o the y-axis agains the results of a full DFT calculation of the 104 atom model system of rubredoxin on the x-axis. Figure reproduced from Ref. 2.

Figure 3.3 shows that taking into account the metal-ligand NBOs and the LNBOs there is a good agreement with DFT calculations of the large 104 atom model systems, which takes the nitrogens into account in the DFT explicitly. Hence, to extract distance information from nitrogen relaxation data it is important to take local unpaired electron spin density into account. In fact the

local nitrogen orbital with unpaired electron spin density can be modelled by a NBO that depends on the spin polarisation pathway[2] or if there is no polarisation pathway by a mixture of a core orbital (CR) and a lone pair (LP).

A local orbital $|\psi_{local}\rangle$ without spin polarisation pathway can be defined as a CR and LP mixture:

$$|\psi_{local}\rangle = \cos\phi\,|CR\rangle + \sin\phi\,|LP\rangle \tag{3.4}$$

where ϕ is the mixture coefficient, which for rubredoxin and the very different blue copper protein plastocyanin is $\approx 86°$, which therefore might be of general applicability for the local orbital occupancy of nitrogens in paramagnetic proteins.[2]

The local unpaired electron spin density ρ_{local} can be determined experimentally by measuring the contact shift $\delta_{con,EXP}$ of the nitrogen. Therefore, it is possible to correct for the PRE effect of the LNBOs to further extract information about the nitrogen-(metal-site) distance. Separating the local orbital ψ_{local} contribution and the metal-ligand NBOs ψ_M one can re-write Equation 3.3 as one now only needs to integrate over the metal-ligand NBOs and the LNBOs with corresponding occupation:

$$r_{eff}^{-6}(\mathbf{r}', \psi_{local}, \delta_{con,EXP})$$
$$\approx \frac{4\pi}{5} \sum_{\nu=-2}^{2} \left| \rho_{local} \langle\psi_{local}| \hat{\mathcal{F}}_2^{\nu} |\psi_{local}\rangle + \langle\psi_M| \hat{\mathcal{F}}_2^{\nu}(\mathbf{r}') |\psi_{local}\rangle \right| \tag{3.5}$$

This approach can be tested by comparison of the effective distance from Equation 3.5 and the crystal structure of rubredoxin. The contact shifts were taken from a previous study[186] and the relaxation measurements of ^{15}N are described elsewhere[2] in detail.

Figure 3.4 clearly shows that a) the NBO approach outperforms the point

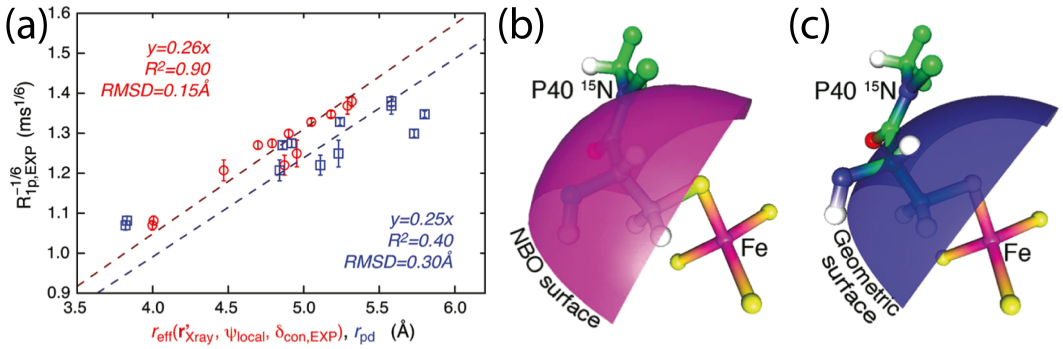


Figure 3.4.: Comparison of the effective electron-^{15}N distances and measured R_{1p} values for rubredoxin. (a) Red circles are the effective distance calculated from the metal-ligand pNBOs, the experimentally measured contact shift and the position from crystal structure. Blue squares are the effective distances calculated with a point dipole approximation; the measured distance in the crystal structure is compared with the measured PRE directly. Vertical bars represent the experimental uncertainty and th dashed line is the best fit. (b) Allowed position of the nitrogen atom of Pro40 calculated from the experimental PRE and contact shift using the LNBO from Equation 3.4 and solving Equation 3.5. (c) Allowed positions for the same nitrogen as in (b) calculated using the point dipole approximation, i.e.$||r_{Fe-N}||^{-6} = R_{1p,EXP}\xi$. Figure reproduced from Ref. 2.

dipole approach in terms or accuracy, as can be seen by the RMSD in panel (a) and that the NBO approach b) yields a correct solution for the position of the backbone nitrogen atom of Pro40 highlighted in panels (b) and (c).

Hence, the simple and intuitive description using NBOs allows to accurately extract distance information for even heavy atoms such as ^{15}N. Therefore, it should be possible to use PRE measurements of nitrogen atoms in isotopically labelled proteins to investigate the structure near a paramagnetic metal-site in solution by means of extracting positional information.

3.1.6. Restrained MD

As described in Section 1.2.2 MD simulations can be biased or restrained with experimental data to e.g. improve the description of the protein structure or dynamics. This is especially interesting for metal-sites as metal ions are usually poorly described in an MD context, where they are mostly modelled as simple beads. For transition metal ions this is of particular importance as a) transition metal ion binding sites often constitute the active site of an enzyme and b) the anisotropic orbital structure of transition metal ions make them hard to model in generally applicable force fields.

To improve the description of the iron-sulphur cluster in rubredoxin and to highlight the capability of the presented NBO approach to yield structural information, a series of MD simulations and restrained MD simulations (using nitrogen contact shift and relaxation data) were preformed.

First the crystal structure of rubredoxin (PDB id 5RXN) was prepared for MD simulations. MD simulations were performed using the Gromacs 4.5.4 package[187] with standard parameters and the AMBER99SB-ILDN[56] force field. The protein was solvated in explicit TIP3P water, and periodic boundary conditions were applied. Thirteen sodium ions were added randomly to neutralise the system. The structure was first energy minimised with 1000 steps of steepest decent gradient minimisation, and subsequently the water and the proton positions were equilibrated by a 100 ps simulation at 298 K with positional restraints applied to all heavy atoms.

To generate structures that are sufficiently diverged from the crystal structure a simulation at 373 K was run to subsequently test whether the NBO distance restraints could improve the structure. The RMSD of the backbone atoms in the metal binding loop (residues 6-11 and residues 39-44) was 1.5 Å after 20 ns of simulation at 373 K, which serves as a starting structure for testing the NBO derived restraints. Furthermore, the iron-sulphur distances from the

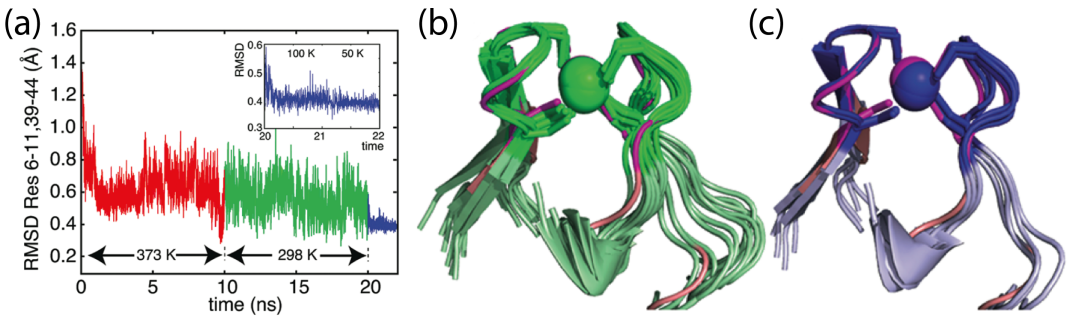


Figure 3.5.: Restrained molecular dynamics of oxidised rubredoxin by incorporation of Fermi contact shifts and ^{15}N paramagnetic relaxation enhancement (PRE) data. (a) Backbone root mean square deviation (RMSD) of the metal binding loops of rubredoxin during the simulation compared to a crystal structure of the protein (PDB id 5RXN). Inset shows a zoom in of the cooling phase of the simulation from 100 K to 50 K. (b) Simulation structures from 10 to 19 ns (green) in 1 ns steps overlaid with the crystal structure (magenta). Residues where paramagnetic NMR data are available are shown with bright colours. Other residues are shown with pale colours. Fe(III) is represented as a sphere. (c) Final structure of the simulations at 22 ns, obtained after cooling to 50 K and overlaid with the crystal structure (magenta). Figure reproduced from Ref. 2.

crystal structure were kept fixed using restraints during production runs to keep the iron-sulphur intact and to not lose the iron atom during the 373 K simulation.

Before testing the improvement upon the implementation of the PRE restraints via the NBO approach or the point dipole approach a 0.5 μs unbiased MD simulations was performed. Here, the RMSD of the residues with available ^{15}N PRE and contact shift measurements does not decrease. This shows that the MD force field is not capable of reaching a structure more similar to the experimentally determined crystal structure.

Now the paramagnetic NMR parameter of rubredoxin were included in the simulation as distance restraints. After each 100 ps of simulation time the pNBOs and LNBOs are transferred from the model system to simulated structure. The

LNBO were then evaluated at three different points along the N-Fe vector and interpolated to find the intersection of the effective distance calculated from the PRE using equation 3.5, which was subsequently converted into a distance restraint. This was realised by using the *genrestr* routine of Gromacs and continuing the run.

As can be seen in Figure 3.5 the inclusion of distance restraints on the starting structures generated by the 373 K run described earlier leads to a significant improvement as measured by the RMSD of the backbone atoms of residues 6-11 and 39-44 (same atom selection as before). The RMSD of the backbone of these residues improves to 0.55 ± 0.1 Å when simulated at 298 K. Cooling the system further improves the structure compared to the crystal structure down to an average RMSD of 0.40 ± 0.03 Å. The average was calculated from three independent runs. Hence, a well-defined structure is obtained for the residues where ^{15}N PRE and contact shift data is available using the implementation of restraints from the NBO approach described above.

While the above shows a clear improvement in the structure of rubredoxin, it might be possible that a less sophisticated approach, i.e. the point dipole approach, yields a similar improvement. Hence, the same procedure was followed with the difference of how the effective distance is calculated (Figure 3.4). Here the inclusion of distance restraints yields an improvement to a RMSD of 0.65 ± 0.02 Å at 298 K, where lowering the temperature to 50 K did only yield a marginal improvement down to a RMSD of 0.63 ± 0.02 Å for the same residues. The RMSD was calculated from two independent trajectories.

In comparison the NBO improves the structural sampling close to the crystal structure significantly, even when compared to an incorporation using the point dipole approximation.

3.1.7. Conclusion

Firstly it was shown that NBOs can be effectively used to describe the unpaired electron spin density of metal-sites in paramagnetic proteins. Approximately 90 % of the electron spin density can be accurately described by only a few NBOs. While using the few metal-ligand NBOs is sufficient to describe and calculate proton PRE and extract distance information from measurement of these, the nitrogen PRE is more complex. However, the addition of a few LNBOs was shown to be sufficient to interpret ^{15}N PRE NMR data and to extract accurate distance information from them. In fact, N-Fe distances can be extracted with an RMSD as low as 0.15 Å, where the point dipole approximation only reaches 0.30 Å (Figure 3.4).

Furthermore, using the NBO approach as described here it is now straightforward to generate accurate distance restraints for MD simulations, where the restraints yield a much improved description of the metal-site as shown by the comparison to the crystal structure. It should however be noted that this does by no means improve the actual description of the metal-site in the modelled system. It merely provides a way to 'correct' the improper description of the metal-site in the used force field by applying experimentally derived restraints. It therefore can not be readily applied to study catalytic reactions of enzymes, such as for HDAC8,[133] as the standard restrictions for MD simulations still apply.

This also means that this method is much faster than QM/MM type simulations as no further DFT type calculations have to be performed after the initial pNBO and LNBO (spin polarisation pathway) setup. Therefore, the presented work adds to the growing list of NMR restraints that can effectively and fast be used to improve the accuracy of structural sampling of MD simulations.

While not part of the presented work, these parameters could also be back-

calculated from a trajectory using the presented approach to test other means of metal-site parameterisation or restraints for MD force fields.

3.2. Measuring Central-Spin Interaction with a Spin-Bath by Pulsed ENDOR

Silicon semiconductors are the building blocks of modern CPU transistors. However, next to the typical transistor-type CPUs the idea of creating quantum computers is a rapidly expanding area of research and development. The concept of quantum computers were first introduced by Yuri Manin in 1980 and Richard Feynman in 1982.[188] Instead of electrical bit-type data, i.e. definite 0 or 1 in electrical transistors, data is stored in quantum-bits (qubits), e.g. spin systems where the entanglement and superposition could be used to effectively store and manipulate data.[189] This type of data storage and manipulation would allow for very different computational algorithms, i.e. quantum computing algorithms such as Simon's algorithm,[190] which has been shown to be superior to any classical probabilistic algorithm.

The efficient realisation of quantum computers depends on creating suitable qubits. While this can in principle be done using different types of quantum systems, e.g. polarised photons, the project I was involved in focused on electron (and to an extend nuclear) spins as a means of forming the qubit. From a theoretical point of view, manipulation of even complex spin systems is very well understood and can readily be done using modern spectrometers. However, most of these techniques are used on bulk magnetised samples, i.e. usually more than 10^{20} spins.

Hence, there are multiple obstacles to tackle in order to build a spin based qubit from a material science perspective as well as from a spectroscopic point of view in terms of manipulation and measurement of single spins, or at least fewer spins. While not the focus of the presented study it should be noted that in terms of detection significant advances have been made by e.g. using single nitrogen-valency (NV) readouts.[191]

Another part of making qubits, is to find a suitable material and spin system,

which is the focus of this project. Silicon-based semiconductors doped with group V donors have been considered to have a great potential for use quantum information processing (QIP),[189] as these systems are amenable to to high fidelity manipulation of spins via magnetic resonance. As will be shown later these systems can in principle be used to store information in the electron spin state as well as the nuclear spins. A major pitfall of these type of this material and electron spins in general is the decoherence time, which especially in the case of electrons is usually very short for memory storage use, i.e. on the order of a hundreds of μs at most.[5]

The decoherence time can be increased using isotopically pure ^{28}Si, which is however very expensive for building any large-scale quantum computers. While it could be argued that in favour of longer decoherence times nuclear spins would be preferable, these have a poor polarisation compared to electron spins at the same external magnetic field and temperature, making the electron spin superior. Moreover, electron spins can be manipulated easily with very fast pulses, i.e. on the order of nanoseconds, whereas nuclear spins need pulses on the microsecond time scale for effective manipulation.

To overcome the problem of short decoherence times of electron spin states in bismuth doped silicon (Si:Bi) I first characterised the superhyperfine (SHF) couplings of the Bi electron spin with adjacent nuclear spins of 4.67 % naturally abundant ^{29}Si in the semiconductor lattice, i.e. the bath spins. These results where then used to calculate the decoherence time due to the dominating spin diffusion at different magnetic fields. The computational part of the work was done by Seto Balian in the Monteiro group (UCL Physics) using coupled cluster expansion (CCE)[192] calculations.

3.2.1. Characterising Si:Bi Using Davies-ENDOR

The aim of this project was to investigate whether an optimal working point (OWP) exists for the Si:Bi material where the SHF couplings of the elec-

tron spin and the nuclear ^{29}Si spins, which are the main contributor towards the decoherence via spin diffusion, vanish and thereby increase the decoherence time. The Si:Bi has an electron spin of $S = 1/2$ and a large nuclear spin of $I = 9/2$ giving rise to 10 different EPR transistions, i.e. 20 different spin states. Moreover, Si:Bi exhibits an extremely large hyperfine constant of $A/2\pi = 1.4754\,\mathrm{GHz}$, which at low magnetic fields ($0.1 - 0.6\,\mathrm{T}$) leads to strong mixing of nuclear and electron transition, in fact allowing forbidden transitions to occur. Due to this strong mixing of spin states it could be shown previously that extrema of the form $df/dB = 0$ exist, where B is the magnetic field and f is the transition frequency in Hz. These studies lead to the hypothesis of OWPs of Si:Bi around these extrema, since line narrowing of EPR transitions and decreased sensitivity to external magnetic field fluctuations were found theoretically.[193]

For Si:Bi the spin Hamiltonian can be written as:[5]

$$\hat{H} = A\hat{I} \cdot \hat{S} + \hat{H}_{Zee} + \hat{H}_{int} + \hat{H}_{bath}, \tag{3.6}$$

where $\hat{I}$ is the Bi nuclear spin operator, $\hat{S}$ is the electron spin operator, $\hat{H}_{Zee}$ is the Zeeman Hamiltonian, $\hat{H}_{bath}$ describes the dipolar interaction of bath spins and $\hat{H}_{int}$ the spin-bath interaction of the electron spin and the nuclear ^{29}Si spins. All terms are described in more detail in Ref. 5, but for the purpose of the presented work the spin-bath Hamiltonian is the most important part. The spin-bath interaction is very important as it has been demonstrated that Si:Bi with isotopically pure ^{29}Si has much longer decoherence times than natural Si Si:Bi. Moreover, it is known that the spin-bath interaction, by the process of spin diffusion, is the major factor that determines the decoherence time of Si:Bi.

In order to investigate the influence of the spin-bath of ^{29}Si spins I first measured the SHF couplings of the Si:Bi material. This data would later be used

to predict where the SHF couplings collapse, effectively decoupling the donor spin from the bath and thereby suppress spin diffusion, increasing the electron spin decoherence time.

3.2.2. ENDOR Spectra of Si:Bi EPR X-Band Transitions

In order to measure the SHF couplings I used Davies pulsed electron nuclear double resonance (pENDOR). A $\pi_{mw} - \tau_1 - \pi_{rf} - \tau_2 - \frac{\pi}{2}_{mw} - \tau_3 - \pi_{mw} - \tau_3 - echo$ pulse sequence was used, where rf denotes radio frequency pulses ($2 - 12\,\mathrm{MHz}$) and mw denotes microwave pulses with a frequency of $\nu = 9.755\,\mathrm{GHz}$ (X-band). The rf pulse frequency was stochastically varied for the range of interest (Figure 3.6) to avoid artefacts that can arise from saturation of 'close' nuclear transitions, which would be the case for sweeping the rf frequency. The delays and pulse length were chosen for the best singal-to-noise: $\pi_{mw} = 256\,\mathrm{ns}$, $\frac{\pi}{2}_{mw} = 128\,\mathrm{ns}$, $\pi_{rf} = 10\,\mu\mathrm{s}$, $\tau_1 = 1\,\mu\mathrm{s}$, $\tau_2 = 3\,\mu\mathrm{s}$, $\tau_2 = 1.5\,\mu\mathrm{s}$ and the shot repetition time was set to 1.3 ms. Presented experiments were recorded at 15 K on a Bruker (Karlsruhe, Germany) E580 EPR spectrometer equipped with pulsed ENDOR accessory E560D-P, a dielectric ring ENDOR resonator EN4118X-MD4, a liquid helium flow cryostat CF935 from Oxford Instruments and radio frequency amplifier ENI A-500W. The donor concentration of the sample was $3 \cdot 10^{15}\,\mathrm{cm}^{-3}$. For the data from Figure 3.6 the external field was directed perpendicular to the [111] plane of the sample crystal. For the rotation pattern (Figure 3.7) of the Si:Bi crystal this orientation corresponds to $0\,^\circ$.

It should be noted that radio frequencies between 0 and 2 MHz were not recorded as the amplifier was not able to give a stable power output for this range, as measured by inspecting the pulses using an oscilloscope. Figure 3.6 shows ENDOR spectra of all 10 transitions. The Figure highlights all visible SHF couplings in different colours and for all SHF couplings that I later identified to be isotropic, i.e. independent of crystal orientation, the value is given. The highest SHF coupling was found to be 11.36 MHz.

As mentioned earlier the eigenstates of $|m_S, m_I\rangle$ become mixed, corresponding to (at most) doublets. The mixed eigenstates can be written as:[5]

$$|\pm, m\rangle = a_m^{\pm} \left|\pm\frac{1}{2}, m \mp \frac{1}{2}\right\rangle + b_m^{\pm} \left|\mp\frac{1}{2}, m \pm \frac{1}{2}\right\rangle, \tag{3.7}$$

where m is an integer such that $m = m_S + m_I$ and $-5 \leq m \leq 5$. It is then straightforward to define a measure of the amount of mixing of the states as:[5]

$$\left|a_m^{\pm}\right|^2 - \left|b_m^{\pm}\right|^2 = \frac{\Omega_m(\omega_0)}{\sqrt{\Omega_m^2(\omega_0) + 25 - m^2}} \equiv \gamma_m(\omega_0), \tag{3.8}$$

where $\Omega_m^2(\omega_0) = m + \frac{\omega_0}{A}(1 + \delta_{Bi})$. Here $\delta_{Bi} = 2.486 \cdot 10^{-4}$ is the ratio of the Bi nuclear gyromagnetic ratio to the electronic one. The derivation of these formula is done elsewhere. Here, it is important to understand the mixing term, which in fact can be directly observed in the spectra shown. Moreover, due to the differences of the Zeeman, hyperfine and SHF coupling the spin-bath term from Equation 3.6 can be approximated as:[5]

$$\hat{H}_{int,l} \approx (\alpha_l \hat{I}_l^z + \beta_l \hat{I}_l^x)\hat{S}^z, \tag{3.9}$$

which describes the coupling of the donor spin to a single ^{29}Si nuclear spin at lattice site l. In principle α_l and β_l can describe any (anisotropic) coupling, i.e. $\alpha_l = [(a_{iso,l} - T_l) + 3T_l^2 \cos^2\vartheta_l]$ and $\beta_l = 3T_l \sin\vartheta_l \cos\vartheta_l$, where $a_{iso,l}$ is the isotropic part and T_l is the anisotropic part of the SHF coupling of the ^{29}Si spin at site l and ϑ_l is the angle between the external magnetic field and the line connecting the donor site and the lattice site l.

Diagonalisation of the resulting Hamiltonian (Equation 3.9), neglecting anisotropic coupling since the vast majority is isotropic, leads to a simple expression for the ENDOR resonance frequencies for a given donor level $|\pm, m\rangle$:

$$\Delta_{iso,l}^{\pm,m}(\omega_0) = \frac{1}{2\pi}\left|-\omega_0\delta_{Si} \pm \left(\frac{a_{iso,l}}{2}\gamma_m(\omega_0)\right)\right| \tag{3.10}$$

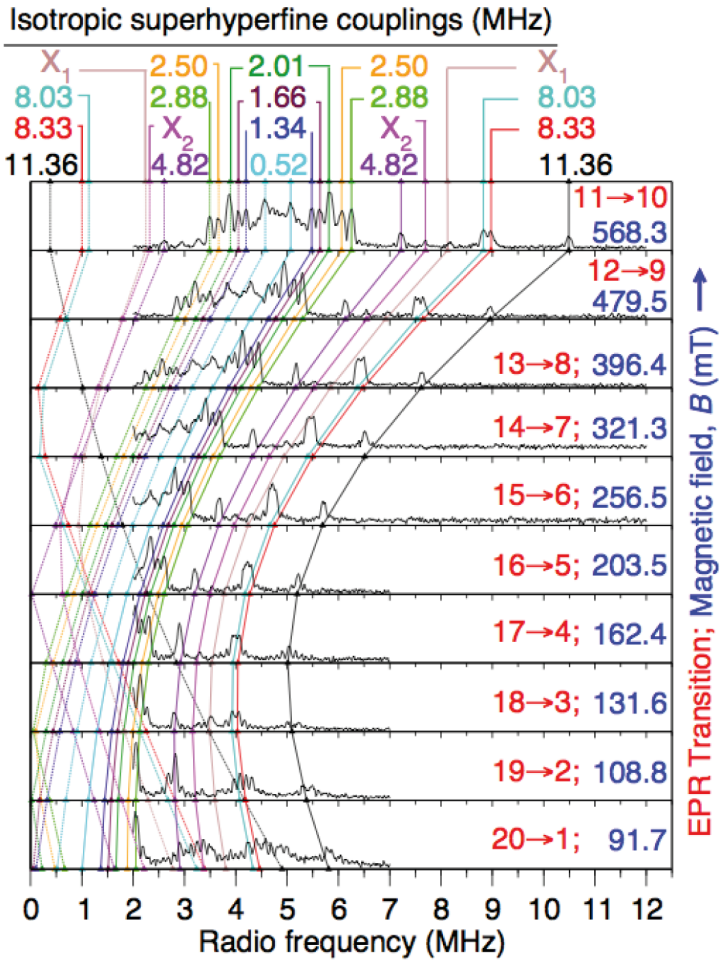


Figure 3.6.: ENDOR of all observable EPR transitions at X-band (9.5 GHz). Eigenstates of the donor levels, and the transitions, are labelled $|1\rangle - |20\rangle$ with increasing energy. SHF coupling values were extracted by fitting equal width gaussians to coupling constants according to Equation 3.10. Figure adapted from Ref. 5.

The resonances in Figure 3.6 have been fitted using equal width gaussians using Equation 3.10 to extract the SHF coupling constants. The coloured lines show the resulting gaussian fit position, which are in excellent agreement. Moreover, it can be seen that the SHF coupling go through a minimum just below $B \simeq 0.2\,\mathrm{T}$, which is a direct result of the mixing γ_m of the involved donor levels.

3.2.3. ENDOR Rotation Pattern of the Si:Bi Crystal

Figure 3.7 shows the ENDOR rotation pattern of the Si:Bi crystal of the $|11\rangle \rightarrow |10\rangle$ transition. To investigate which of the peaks resolved are anisotropic the crystal was rotated in 10° increments. Isotropic peaks should stay at the same position and anisotropic peaks should shift depending on the rotation.

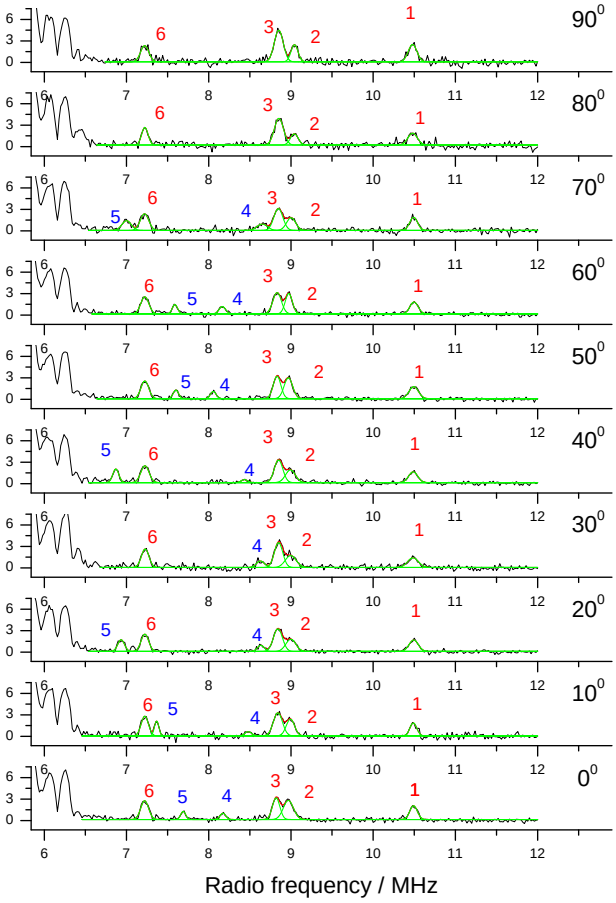


Figure 3.7.: ENDOR rotation patter of the $|11\rangle \rightarrow |10\rangle$ transition. The spectra are zoomed in to the large couplings, i.e. larger than the ^{29}Si Larmor frequency. Black lines show the experimental data and the green lines a series of fitted gaussians, for which the parameters are shown in Figure 3.8. Six peaks are highlighted, where the isotropic ones are labelled in red and anisotropic ones in blue.

From Figure 3.7 it becomes immediately clear that the majority of the peaks are indeed isotropic. I could only identify three SHF couplings that were anisotropic, where one is not resolved in Figures 3.7 and 3.6. Peaks 4 and 5 are the ones easily resolved that are clearly anisotropic. This can be seen very well from Figure 3.8. These two peaks are the only ones exhibiting a clear dependence on the orientation of the crystal. Apart from only observing very few anisotropic couplings, the peaks of these couplings are very weak compared to other couplings, which can be seen from the 'Area' column in Figure 3.8.

There is, to the best of my knowledge, only one other study where electron spin echo envelope modulation (ESEEM) is used to identify a single anisotropic coupling of a ^{29}Si in the E-shell (next neighbour of the donor).[194]

3.2.4. Proposing an Optimal Working Point For Si:Bi

The above established a basis for the investigation of the Si:Bi material as a) SHF couplings of the material have been measured and b) the crystal pattern shows that isotropic SHF couplings are dominant with only three very weak anisotropic couplings observable. Hence, in a model the anisotropic could be neglected in good approximation. However, it could be shown that even considering anisotropic SHF couplings yield the same conclusion in terms of spin diffusion suppression at the OWP.[5]

Seeing how the SHF couplings behave, i.e. depending on the mixing of the donor levels, the possibility of a collapse of these couplings could be envisioned as Equation 3.10 tends to the ^{29}Si Larmor frequency for $\gamma_m = 0$ (this is also the case for anisotropic couplings). However, γ_m is not zero for two donor level at the same magnetic field. Yet, this is not necessary, i.e. for the $m = -3, -4$ transition an OWP exists at $B = 188\,\mathrm{mT}$, when:

$$\gamma_{-3}(\omega_0) + \gamma_{-4}(\omega_0) - \frac{2\delta_{Bi}}{1+\delta_{Bi}} = 0 \tag{3.11}$$

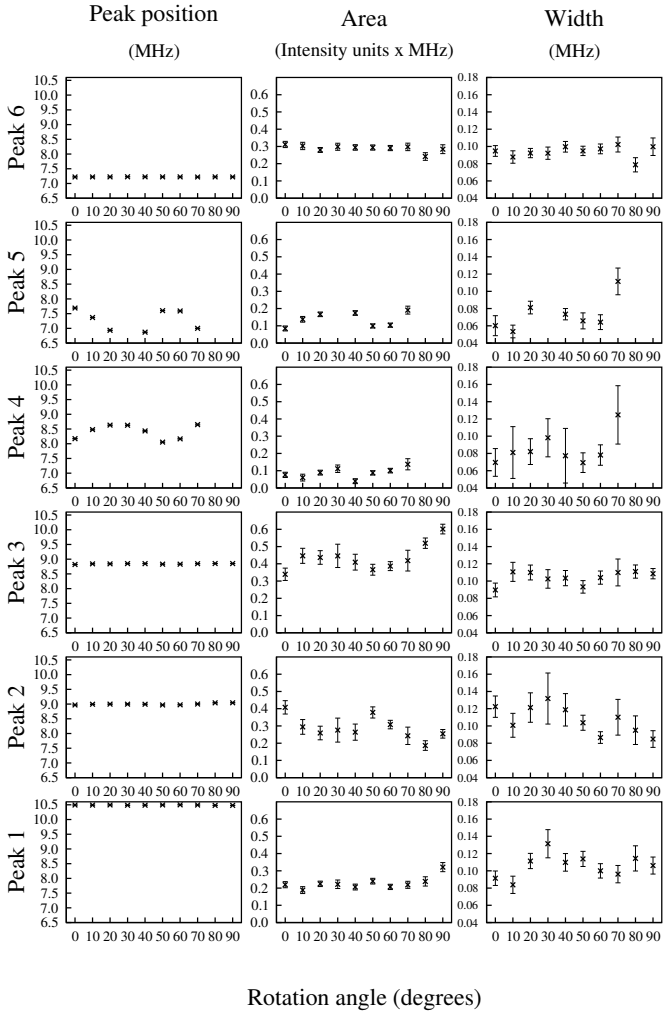


Figure 3.8.: Summary of the crystal rotation pattern fits using gaussians. Results per peak are plotted, the first column reports on the peak position (in MHz) over the rotation angle, the second column reports the integral of the fitted peak and the third column reports on the width. The resulting fits are illustrated in Figure 3.7 by overlaying the fits with the experimental data.

At this point the donor effectively decouples from the spin-bath, which should lead to suppression of decoherence via this the spin diffusion process. To analyse this, Seto Balian applied coupled cluster expansion (CCE) calculations to evaluate the decoherence time around the the OWP.

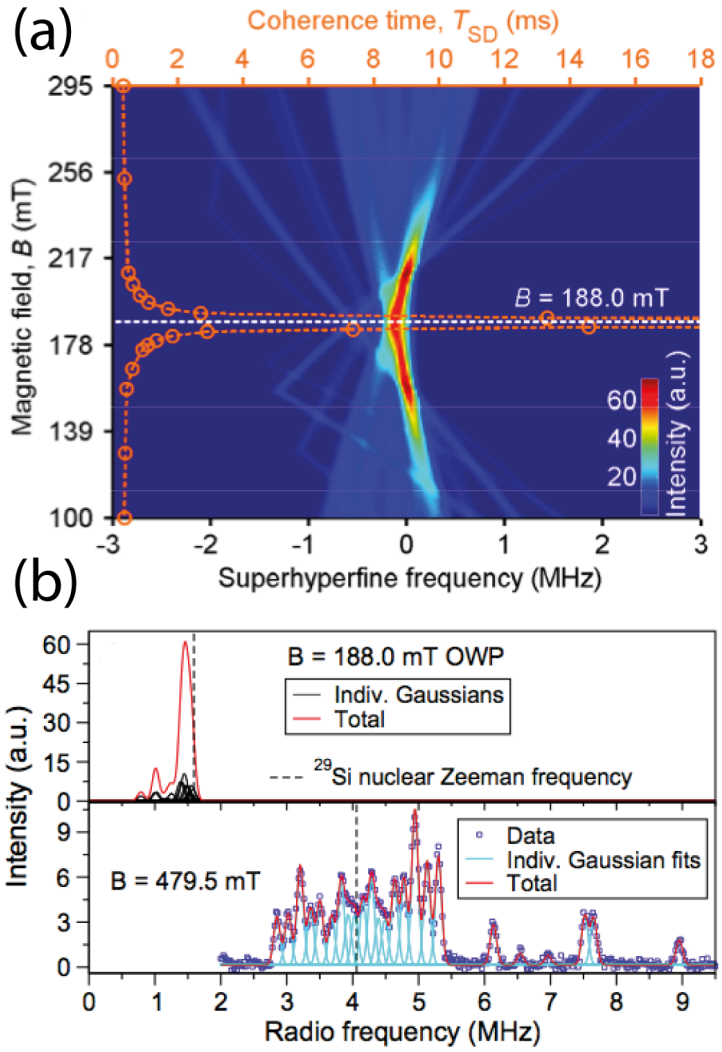


Figure 3.9.: Calculated ENDOR spectra and decoherence time of the $|12\rangle \rightarrow |9\rangle$ transition at 188.8 mT. (a) Color plot of ENDOR resonances in the magnetic field-SHF frequency space. The overlaid orange line is the spin decoherence time calculated using the CCE at specific fields. (b) Upper panel shows the ENDOR spectrum of the $|12\rangle \rightarrow |9\rangle$ at the OWP and the lower panel shows the measured ENDOR spectrum at X-band of the same transition.

In brief, in the CCE different spatial configurations of spins are simulated taking into account only dipolar couplings, as these are the perturbative ones with a few kHz. The problem of calculating the ensemble decoherence time is therefore split into clusters of defined spatial spin configuration. 100 spatial

configurations have been calculated using at most 2 bath spins, which lead to a converged CCE. In each cluster ^{29}Si spins were separated by up to the third nearest neighbour site in a cubic Si lattice with a 160 Å side length.

The dipolar interaction between the bath spins are modelled by simple Ising ($\hat{I}_1^z \hat{I}_2^z$) and flip-flop ($\hat{I}_1^+ \hat{I}_2^- + \hat{I}_1^- \hat{I}_2^+$) terms. The resulting decay curves for the calculated Spin-Hahn echo were fitted to a simple exponential function to yield the relaxation times reported in Figure 3.9.

The results of these calculation are surprising as the decoherence time increases asymptotically very sharply around the OWP. Figure 3.9 (a) shows the overall behaviour of the material: the SHF couplings of the $|12\rangle \rightarrow |9\rangle$ transition collapse at the OWP and the decoherence time increases. Hence, the decoupling of the donor from the spin bath leads to suppression of spin diffusion causing the increase in decoherence time.

The collapse of the SHF couplings is illustrated as well in Figure 3.9 (b), where the lower panel shows the measured ENDOR spectrum of the $|12\rangle \rightarrow |9\rangle$ transition and the upper panel shows the same transition at the OWP. The OWP illustrated here is not the only one, in fact there are several for this material in the frequency range of $5 - 7.5\,\mathrm{GHz}$. Furthermore, it has been suggested that spin-bath decoupling also leads to suppression of relaxation arising from a donor bath, which has however not been investigated here.

3.2.5. Conclusions

The presented study is the first reporting on the (dominantly) isotropic SHF couplings of Si:Bi, which arise from the ^{29}Si impurities. Further, it was demonstrated that these couplings are suppressed at the OWPs. Finally, the calculating the spin decoherence time of the donor spin as a many-body problem showed very sharp divergence of the decoherence time at the OWP of the $|12\rangle \rightarrow |9\rangle$ transition.

Therefore, this study strongly motivates further experimental investigations around the suggested OWP of Si:Bi. In fact, the possible implications of the presented results could be of great importance, as natural Si:Bi is much cheaper than isotopically enriched Si:Bi to produce suitable qubits. Finally, the presented work also serves as proof-of-principle, as it makes obvious that one is able to simultaneously manipulate the electron and nuclear spins involved using rapid pulses, which has been suggested by the Monteiro group.[193]

3.3. Measuring In Vivo Iron Levels in C. Elegans

This project was carried out as part of ongoing work in the group of the Gems group (UCL, GEE). I have measured and analysed all EPR spectra.

The phenomenon of ageing of cells and organisms in general is well-known, but is a somewhat poorly understood process in life-sciences. In fact the quote from Erwin Schrödinger in the preface of this thesis underlines how fundamental this process is.

There are different hypotheses of how ageing works on a mechanistic level.[195] One of the most prominent theories is that biological materials are deteriorated over time. Specifically the accumulation of damage caused by reactive oxygen species (ROS) has been thought to be responsible for ageing of cells.[196] This theory has been followed up in many different animal models, which has however not lead to conclusive answer.[197]

Oxidative damage in the short-lived nematodes *Caenorhabditis elegans* has been studied, which did not support the ROS ageing theory.[198] In fact, it was shown that deletion of *sod-2*, a gene that encodes the major mitochondrial dismutase SOD, which removes the superoxide free radical, even leads to an increase of the life-span, rather than a decreasing it.[198]

Taken together, the literature findings do raise doubts about the validity of the oxidative damage theory.[199]

3.3.1. Iron as a Generator of Oxidative Damage

Iron is a key metal in many biological processes, its uses include the transport of oxygen and the mitochondrial energy metabolism. This metal has different oxidation states, where ferric(II) and ferrous(III) forms are present *in vivo*. A lot of iron atoms are complexed, but there is a pool of not-complexed 'free' iron. This pool of 'free' iron can be toxic, partly due its involvement in the Fenton reaction; Fe(II) is oxidised by hydrogen peroxide (H_2O_2), which pro-

duces the reactive hydroxyl radical $OH^{\bullet}$. Other metal ions can catalyse the Fenton reaction, but iron is thought to be the major contributor towards ROS production.[200]

Hence, controlling iron might also be key to ageing. In fact, iron chelation has been speculated as a treatment to protect against ageing, as it could be shown that this does decrease oxidative damage.[201] There are iron storage proteins, so-called ferritins, which bind 'free' Fe(II) making it unavailable for the Fenton reaction. Therefore, ferritins can be seen as effective antioxidants. In *C. elegans* there are two ferritins, FTN-1 and FTN-2.[202] The expression of the *ftn-1* gene has also been shown to be increased in long-lived *daf-2* mutants, which might give a clue to whether iron levels are important for resistance against oxidative stress or the life-span of the organism.

Investigating whether and how 'free' iron levels might contribute towards oxidative stress resistance and life-span of *C. elegans* was the focus of the presented study. The above illustrates that there are several clues that this might be the case and gives multiple ways to modify the 'free' iron level present in the organism: a) supplementation/chelation of 'free' iron, b) mutation/knockdown of *ftn-1/ftn-2* genes and c) change expression level of *ftn-1* gene via mutating *daf-2*.

Using these approaches and monitoring the 'free' iron levels, oxidative damage and life span of the nematodes was therefore thought to give a more detailed answer to how the Fenton reaction contributes to ageing, at least in *C. elegans*. To achieve this type of monitoring a multitude of experiments were employed. Except the EPR experiments these techniques will only be outlined briefly, a more detailed description can be found in Ref.[3] and references therein.

3.3.2. Sample Preparation

All samples were prepared by Sara Valentini in the Gems laboratory (UCL, GEE). In general nematodes (different strains) were raised in Petri dishes using nematode growth media (NGM) with *E. coli* OP50 as food source. Supplementation of iron was achieved having the nematodes on NGM plates that were supplemented with ferric ammonium citrate (FAC - $C_6H_8O_7nFe_nH_3N$). For iron chelation the Desferal (DF - $C_{25}H_{48}N_6O_8$) chelator was used. RNA-mediated interference was achieved by *E. coli* feeding.

Protein oxidation was assessed using an Oxyblot kit from Millipore according to the manufacturers protocol. The life-span of worms was monitored by counting alive and dead worms (dead worms were removed from the test plate) every second day, using the L4 stage as time 0. Here, the test plates contained the supplemented FAC/DF/RNAi etc. For the oxidative stress assay worms were scored for survival on test plates containing *tert*-butyl hydroperoxide.

3.3.3. Using EPR to Measure Free Iron in C. Elegans

Monitoring the 'free' iron level of the nematodes was achieved using EPR experiments of worm samples, which was my part of the project. In order to measure 'free' iron I used continuous wave EPR (cw-EPR), which means that in contrast to pulsed experiments such as in Section 3.2.5 there is a continuous 9.5 GHz microwave and the magnetic field is swept across the region of interest. Of course, frequency and field are interconvertable, but in the field of EPR spectroscopy different spin species are often described by what is called the g-tensor, which describes the resonances along three principal components of the molecular frame. In general, the g-tensor (or in the isotropic case g-value) resonance relation is expressed as $h\nu = g\mu_0 B_0$, where h is Planck's constant, ν is the microwave frequency, μ_0 is the electronic Bohr magneton.

Hence, the g-tensor reports on the unique electronic structure of the spin sys-

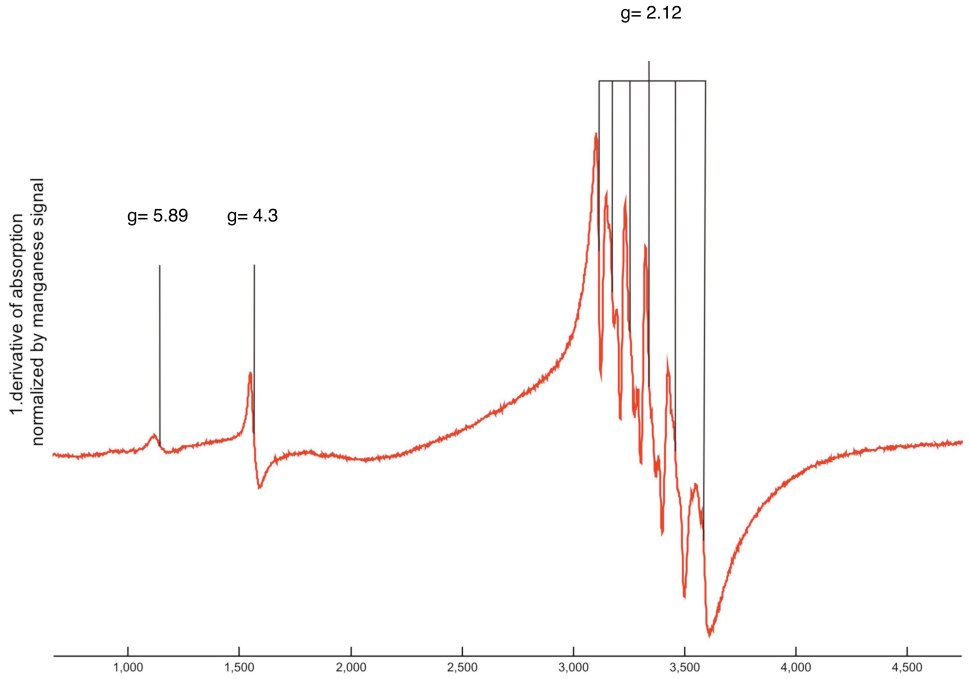


Figure 3.10.: Continuous wave EPR spectrum of wild type *C. elegans* at 10 K. g-values of different signals are highlighted: At $g = 5.89$ a weak heme signal is observed, at $g = 4.3$ the 'free' Fe(III) is observed and at $g = 2.12$ the manganese signal is detected. The manganese signal was used in all subsequent comparative measurements as a means of normalising the EPR signals, assuming that the manganese content per worm is constant. Figure adapted from Ref. 3

tem, which is of course very different for different oxidation states or different coordination of Fe. Most important for the presented work is high-spin Fe(III), which has a sharp component of the g-tensor of 4.3, which can be readily observed in an cw-EPR spectrum. In order to get high-spin Fe(III) a chelator has been used as has been done in a previous similar study.[203]

Age synchronised worms (young adults) were raised on agar plates and collected minimal medium buffer (M9). Subsequently worms were washed to remove any supplement access, e.g. additional FAC or DF. Afterwards the samples were centrifuged for 3 min at 4 krpm at 4 °C. After removing the supernatant

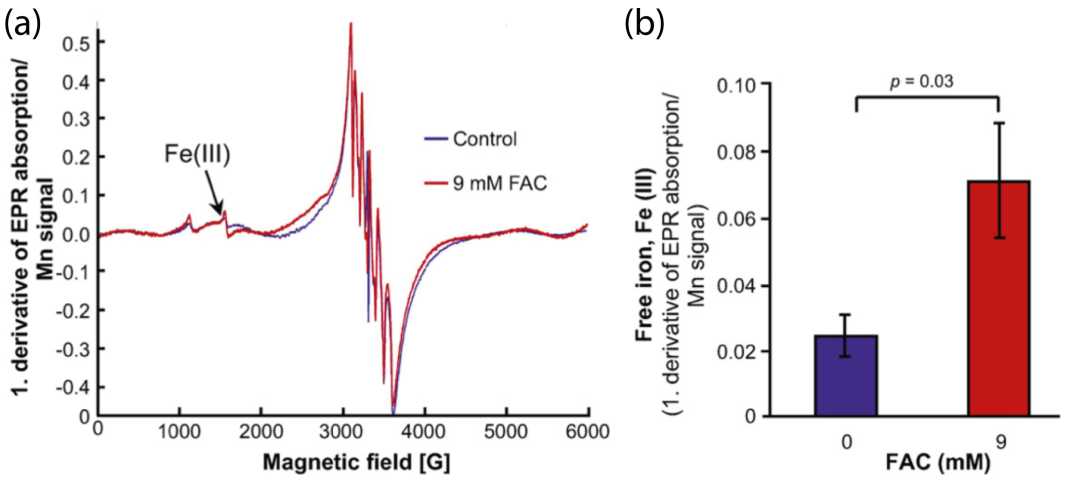


Figure 3.11.: Effect of iron supplementation on *C. elegans*. (a) Continuous wave electron paramagnetic resonance (EPR) spectrum of flash frozen *C. elegans* fed with 9 mM FAC (ferric ammonium citrate) in red and a control in blue. Supplementation with FAC significantly increases the amount of free iron in the organism, as shown in (b). Figure reproduced from Ref. 3.

the worms were resuspended in 15% v/v glycerol. Finally the iron chelator deferoxamine was added to a final concentration of 2 mM. The samples were then allowed to incubate for 15 min while the samples were transferred into 5 mm EPR quartz tubes using 310 mm Pasteur pipettes. After the incubation time the EPR tubes were left on ice to let the worms accumulate on the bottom of the tube, after which the supernatant was removed with a pipette and the samples were flash frozen in liquid N_2. If the sample was not measured right away the sample was kept in liquid N_2 until its use.

EPR measurements were performed at X-band (9.5 GHz) on a Bruker (Karlsruhe, Germany) EMXPlus spectrometer. The spectrometer was equipped with a super high Q 4112SHQE resonator, which had an ESR900 cryostat equipped to allow measurements at a temperature of 10 K. This is necessary as the relaxation time of high-spin Fe(III) is too small at higher temperatures to be measured directly. For all measurements the magnetic field was swept from 0

to 6000 Gauss using a 2 mW microwave power and a modulation amplitude of 10 Gauss with a frequency of 100 kHz.

As expected a high-spin rhombic ferric iron was detected at $g = 4.3$ (Figure 3.10). Since there is Fe(II) present as well worms were supplemented with DF to convert Fe(II) to observable Fe(III). Fe(II) is too broad to be detected and using DF allows a measure of the overall 'free' iron pool, i.e. Fe(II) and Fe(III).

3.3.4. Manipulating the Available 'Free' Iron Does Not Alter Life-Span

As there are multiple ways of altering the 'free' iron pool in *C. elegans* it first needed to be established that supplementation of FAC or knockdown of genes (*ftn-1/2*) leads to the increase of available Fe(II) and Fe(III). Hence, this was the first objective. Figure 3.11 shows that supplementation of 9 mM FAC did in fact yield an increased 'free' iron pool, which the peak at $g = 4.3$ in the EPR spectrum reports on (Figure 3.11 (a)).

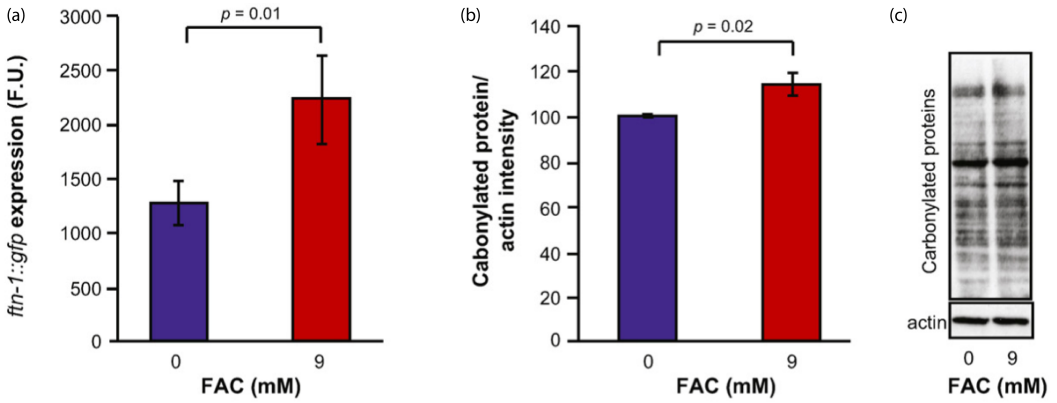


Figure 3.12.: Effect of iron supplementation on protein carbonylation in *C. elegans* . (a) Expression level of *ftn-1* when supplementing iron in the form of FAC. (b) Statistical analysis of three biological replicas of proteins carbonylation, normalised by actin immunoreactivity. (c) Oxyblot from one of the replicas analysed for the plot of (b). Figure adapted from Ref. 3.

Hence, the 'free' iron pool can be influenced by feeding the worms with iron in

the form of FAC. We could measure this at different concentrations, but in general high concentrations of several mM are generally not physiological. When feeding the worm with up to 9 mM FAC the 'free' iron level in the organism increased, but the life span did not decrease. This seemed very surprising since 9 mM supplementation seems like rather high supplementation.

Moreover, in the case of over-expression of *ftn-1/2* or introduction of mutations in *daf-2* no significant changes of the Fe(III) signal could be observed with EPR.[3] Variations between biological replicas of these samples were of similar size as the variation of the baseline signal.

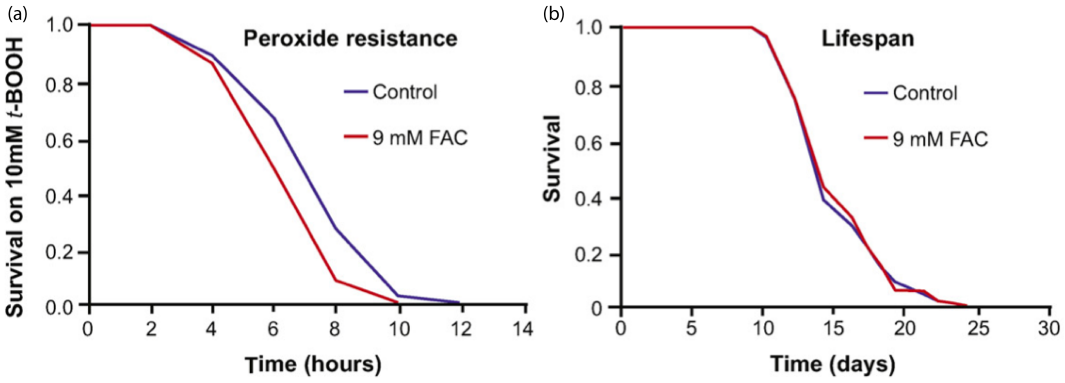


Figure 3.13.: (a) Peroxide resistance test comparing control worms (blue) with 9 mM FAC fed worms (red). (b) Life span analysis of control worms (blue) and 9 mM FAC fed worms (red). Figure adapted from Ref. 3.

While this is somewhat disappointing from an EPR perspective, which is in fact a very sensitive tool, it hints that the effects and general physiological variation of the 'free' iron levels from *ftn-1/2* knock-down or over-expression are very subtle and that there might in fact be other processes at play. Once it could be established that the free iron levels can be altered, at least in the from of FAC supplementation Valentini et al. investigated what the influence on oxidative stress and life span was. Figure 3.12 highlights the influence of iron supplementation on the carbonylation, i.e. the oxidation, of proteins within *C. elegans*. Figure 3.12 (a) shows that supplementation of iron does lead to

an increase of *ftn-1*, which is somewhat expected as the gene is thought to be coding for an iron storage protein. Hence, an increase in the expression of this gene would be an expected response of the worms to tackle the elevated iron level present under FAC feeding conditions.

Moreover, Figure 3.12 (b-c) show that supplementation of iron does also lead to increased oxidative damage as more carbonylated proteins are present overall. All in all these results seem to agree with the prominent hypothesis that the Fenton reaction, an increased 'free' iron level, can be associated with more oxidative damage. Hence, the results were in agreement with what has been suggested by others.[201]

However, all these results are only really relevant when set into context with experiments that test the life span of the organism under the same conditions. Here Sara Valentini measured the peroxide resistance and life span of the worms as described earlier. Figure 3.13 (a) shows that, as expected, the iron fed worms are less resistant to peroxide, which is expected as the peroxide production should be elevated by the increased 'free' iron pool already. However, as stated earlier, the life span is surprisingly not changing.

Hence, a moderate iron supplementation seems to lead to an increase in oxidative damage, but does not decrease life span, which puts the direct hypothesised causality in question.

To test this further Sara carried out experiments to investigate the influence of *ftn-1* over expression and iron chelation on the oxidative damage of proteins and the life span of the organism. Both, the over expression of *ftn-1* and chelation of iron using feeding with 100 μM DF, are more subtle than the moderate supplementation using FAC where the change in the 'free' iron pool could not be measured using EPR. In both cases it was found that the amount of oxidative damage did change, but the life span did not. Hence, while there seems to be consistent effect of 'free' iron levels increasing oxidative damage, an effect on life span of any of the two could not be observed.

3.3.5. Conclusions

Firstly, using EPR it could be shown that it is in fact possible to increase the 'free' iron levels *C. elegans* using FAC supplementation. A high dose of iron supplementation ($> 9\,\mathrm{mM}$) via FAC did decrease the oxidative damage of proteins of the worms and decreases the life span. However, a moderate change in 'free' iron levels ($\approx 200\%$ using $9\,\mathrm{mM}$ FAC)[3] shows oxidative damage coherent with the damage observed in high dose supplementation, but no decrease in life span. Knock-down and over expression of the gene *ftn-1*, which codes for an iron storage protein, changed the 'free' iron level below the significant detection limit of EPR when compared across biological replicas. Here, a consistent oxidative damage response was observed, but again no change of life span.

All in all, the presented results put the hypothesis about the involvement of the Fenton reaction and ROS production and its implication in ageing in question. At least under standard lab culture conditions the change of 'free' iron levels does not seem to change the life span according to the oxidative damage leading to ageing dogma.

Furthermore, it was shown that EPR can be used to asses supplemented 'free' iron levels in *C. elegans*, when using moderate and high dose supplementation. By using the the Mn signal as a way of normalising the spectra a new step was introduced to remove errors that can arise from different biological sample sizes. At least when assuming that the Mn content of the worms is not effected by the different feeding, knock-downs or over expression of the investigated genes.

4. Conclusions

The results presented in Section 2 and 3 are briefly summarised and set in context with the overall aim of the thesis. This includes my investigations of the dynamic interactions of the L1 and L2 loop, the formation of a transient tunnel that could form a product release channel in HDAC8, the NBO approach to use in restrained MD to better describe metalloproteins, the investigation of quantum computing material and the measurement of free iron pools in C. elegans in relation to their life span.

4.1. L1:L2 Loop Interactions and Dynamics

Sections 2.1 to 2.4 describe my investigations regarding the dynamic L1 and L2 loops of HDAC8. Based on molecular dynamics simulations I proposed a model in which dynamic interactions of the L1 and L2 loops steer the behaviour of the binding rail of HDAC8. Specifically, electrostatic interactions of the two loops, Lys33 in the L1 loop and Asp87-89 in the L2 loop, seem to able to change the conformation of functionally important residues, i.e. the binding rail residues Tyr100 and Asp101.

Informed by these simulations I designed single point mutations that interfere with the L1:L2 interaction. The activity assay results presented in Section 2.2 suggest that the interaction of the loops is present in solution and does promote the enzymatic activity of HDAC8. To study the enzyme's activity I designed a non-invasive NMR based assay (Section 2.3), which is now used as a general tool to obtain kinetic data of PTM erasing complexes. The tool can in principle

also be used to study the modification of peptides in real-time, not just the removal of e.g. methyl or acetate groups.

I further investigated the dynamic L1:L2 interactions using advanced NMR tools, i.e. deuterium relaxation experiments on $^{13}CH_2D$ labelled Ile side-chains (Section 2.4). Using this technique I was able to show that the Lys33Glu mutation (which should disturb the L1:L2 interaction) does lead to specific structural changes, most notably of Ile93, which is located in the L2 loop. Hence, the data again points towards a L1:L2 interaction. However, the deuterium relaxation data presented in Section 2.4 did not provide enough detail to observe possible changes in dynamics of loops upon mutating Lys33.

The results of this study created a first model for the mechanistic involvement of the L1 and L2 loop dynamics and interactions in the catalytic cycle of HDAC8. Whether or not this is related to substrate binding or merely substrate specificity remains to be seen. In any case, the changes in activity observed in Section 2.2 are in the same ballpark as other means of down-regulation of the enzyme that have been linked to *in vivo* down-regulation[131] or disease missense mutations.[121]

Specifically, the activity changes observed upon mutating residues in the L1 and L2 loop are 2 to 50-fold. The down-regulation by phosphorylation of Ser39 or the introduction of a mimicking mutant (Ser39Glu) is 2 to 6-fold,[131] the down-regulation of a second zinc binding at the L2 loop is 2-fold[127,129] and the down-regulation of many CdLS mutations are 2 to 5-fold[121,123] *in vitro*. Thus, the proposed mechanism should be regarded as physiologically relevant.

Hence, the mechanism I propose builds a platform to rationalise the influence of the L1 and L2 dynamics on the enzymatic activity of HDAC8 and possibly other HDACs. In fact, the described mechanism is capable of directly explaining the down-regulation of the second zinc binding site. In the structure where this site is occupied by zinc it is clear that the Asp87-89 residues are not available to form electrostatic interactions with the L1 loop.

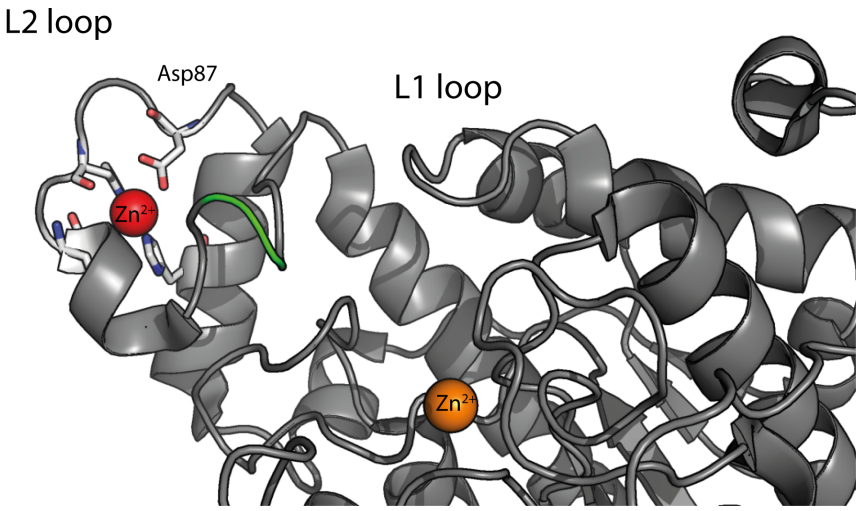


Figure 4.1.: Crystal structure of HDAC8 (PDB code 3F07) [127] shown in grey cartoon representation. Catalytic zinc is shown as an orange sphere. The second zinc is shown as a red sphere. Residues coordinating the second zinc in the L2 loop are shown in stick representation and Asp87 is labelled. The binding rail residues Tyr100 and Asp101 are coloured in green.

Figure 4.1 shows this structure (PDB code 30F7; chain A), where the Asp87 is highlighted as it coordinates the zinc. In fact, the data agrees with a model where L1:L2 interactions lead to increased enzymatic activity. In the structure shown in Figure 4.1 the L1:L2 interaction is hindered by the zinc bound and the mutations I introduced interfere with the electrostatic interactions of the loops. Both events decrease the enzyme's activity.

Interestingly, the regions that I have identified to be important for catalytic activity of HDAC8 were also very recently shown to be important in HDAC3, where two binding partners (NCOR2 and inositol tetraphosphate) are needed for activation (Figure 4.2). [130] In the recently published crystal structure of HDAC3, the activating binding partner (NCOR2) clamps the interface adjacent to the L1 and L2 loops. This clamping yields a structure of HDAC3 similar to that of HDAC8, possibly allowing for a similar L1:L2 interaction and allosteric mechanism. A similar feature can also be observed in HDAC2, where a recent

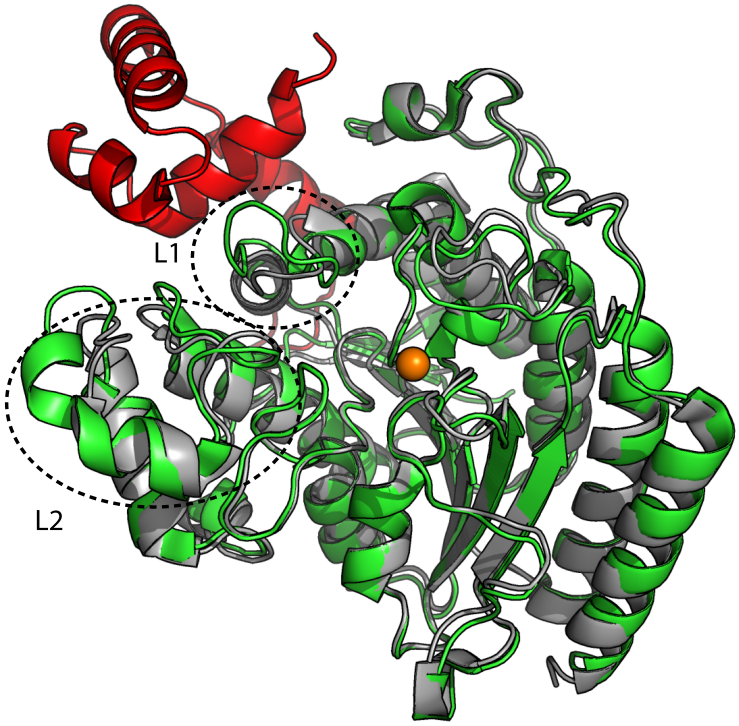


Figure 4.2.: Overlay of the crystal structures of HDAC8 as grey cartoon (PDB code 2V5W) and HDAC3 as green cartoon (PDB code 4A69) bound to the nuclear receptor co-repressor 2 (NCOR2) in red.[130] The catalytic Zn^{2+} is shown as an orange sphere and L1 and L2 loops are highlighted.

crystal structure revealed a more helical L2 loop, but a longer L1 loop, again allowing for inter loop interactions.[204]

My findings regarding the L1 and L2 loop dynamics and interactions seem to fit within the current view of the field, as the proposed mechanism is capable of rationalising observations of other groups. Whether or not the loop interactions do steer the binding rail and thereby tune the activity of HDAC8 remains to be seen.

4.2. Transient Product Release Tunnel

In Section 2.5 I presented my study of a product release tunnel that has been hypothesised by many groups.[19,103,126,174,175] To investigate this from a MD perspective I wrote programmes to analyse water transport between the active site of HDAC8 and the bulk solvent via the internal cavity of the enzyme. I was able to show that the water molecules in the simulation in fact find a specific tunnel that is gated by His42 and Tyr18.

This observation is in line with early proposals of a product release tunnel by Vannini et al.[103] and very recent studies of HDAC-1 and -2.[175] While only water is observed it seems very likely that the product acetate is able to use the same tunnel to be released into the bulk solvent.

Hence, my studies form the first detailed characterisation of such a tunnel in HDAC8. Moreover, the observed gating mechanism accounts for at least one CdLS missense mutation, i.e. Ile19Ser, as this mutation is directly next to the gating residue Tyr18.

Furthermore, the analysis strategy and software can be used generally to track water molecules in transiently forming pores or tunnels and could hence be used to study other similar problems.

4.3. Remarks on the HDAC8 Studies

My investigations show that HDAC8 has structurally dynamic features that seem to have an impact on the catalytic activity of this enzyme. The strategy followed in this thesis, i.e. deriving a working hypothesis from MD simulations with subsequent experimental testing, is general and is capable of aiding deciphering the influence of structural dynamics on protein function.

This is true for the investigations of the L1:L2 loop interactions as well as the transient product release tunnel. MD simulations can be a wonderful tool to

'visualise' protein dynamics on an atomistic scale. The same is true for the use of NMR as it can capture temporal and spatial dynamics of e.g. bond vectors. The potential insight one can achieve with the combination of both techniques is enormous.

4.3.1. Outlook of HDAC8 Studies

The presented results regarding HDAC8 yield a first glimpse at how dynamic features of the enzyme might be important for its function, which could be followed up in the future. How and whether the L1:L2 loop interactions change the conformational sampling for the binding rail could be investigated with biased MD computationally and using NMR, as shown in Section 2.4.

Using the Ile assignment that has been established during my time as PhD student allows for probing the L1 loop via Ile34 and the L2 loop via Ile 94, which enables the use of methyl NMR techniques. While my NMR studies of HDA8 where not conclusive, further investigations will certainly yield more detailed insights into the structural dynamics of HDAC8 given the progress that has been made in the Hansen lab.

Another possibility apart from methyl NMR is the improvement of the established NMR backbone assignment of HDAC8 and later use of chemical shift biased MD to study the dynamics of the protein.

Another important aspect of my work turned out to be the development of the NMR based p53 assay. While it is a simple approach as such it really comes to its own when used on systems that entail different PTM regulators, e.g. a demethylase and a deacetylase as in the CoREST complex as briefly described in Section 2.3. In fact, during my time as a PhD student we established a collaboration where we make use of this assay to study the possible interdependence of two PTM regulators within a complex.

Finally, the tunnel hypothesis could also be explored further with e.g. a simple

single mutation study. Here, a Tyr18Phe mutation should interfere with the gating mechanism proposed in Section 2.5 by only removing the OH group that in the MD forms hydrogen bonds with the water in transit. Furthermore, this could be explored by e.g. enhanced sampling or steered MD. Especially steered MD seems promising, i.e. repeatedly pushing the product acetate through the proposed tunnel at constant velocity - making use of the Jarzynski equality[205] to a) validate (or identify) the tunnel and b) characterise the energy barrier of transport.

All in all, there are plenty of promising opportunities to take the presented research regarding HDAC8 further.

4.4. NBO Approach

The presented NBO approach to describe delocalised unpaired electron spin density in paramagnetic proteins has proven powerful to interpret even rather complex heavy atom (in case of the presented work ^{15}N) NMR PRE data. The investigations presented in Section 3.1 show that besides extracting information for hydrogen-(metal-site) distances using only a few NBOs, even nitrogen-(metal-site) distance information can be extracted from nitrogen PRE and contact shift data by simply adding a few LNBOs to describe the local orbital structure with unpaired electron spin occupation.

Furthermore, I was able to perform MD simulations to show that implementing these extracted distance in terms of restraints significantly improves the accuracy of the protein structure description in the simulations. Here, the developed NBO approach clearly outperforms the typical point dipole approximation in terms of accuracy.

The work performed in this project adds to the development of merging experiment and simulation as only with sophisticated, yet fast modelling an efficient interpretation of experimental data is possible. Therefore, it adds to the list

of methods, such as chemical shift based MD,[206,207] that effectively integrate experimental data directly into simulations. In the presented case the modelling using NBOs is a) intuitive on a pictorial level and b) simple in terms of covering the unpaired electron density in only a few orbitals, which essentially speeds up the integration necessary for calculating the effective distance during MD simulations.

As stated earlier in Section 3.1 this method could also be used for comparison of back calculated data from MD simulations to test other means of improved metal description, such as the recently published non-bonded dummy model for transition metal ions.[208]

4.5. OWP in Si:Bi

While the subject quantum computing might be seen as out of touch with the biological theme of the thesis, it is another example of combining computation and experiment. Here, I performed many experiments to characterise a rather novel material, Bismuth doped Silicon. After characterising the material using pulsed ENDOR Seto Balian used this characterisation to investigate the behaviour of the material under different conditions, i.e. different fields / frequencies.

With the combination of the experimental characterisation and the CCE calculations it became clear that the investigated Si:Bi material has some very unique properties. In fact, the simulations suggest that the SHF couplings can vanish at a certain field for a certain transition. The result of this is that the decoherence time, which is dominated by spin diffusion, basically vanishes as the donor spin basically uncouples at his point, which can then be called an optimal working point (OWP).

While the proposed existence of this point is of great interest by itself, it is particularly interesting since the only other way of getting similar electron

decoherence time in Si:Bi is using isotopically pure (i.e. ^{28}Si), which is of course very expensive. Here, the investigations showed that a very long electron spin decoherence time should be achievable with normal Si, which would make the Si:Bi material much cheaper to produce.

Again, it was the combination of experiment and simulation that lead to an advance in the field of studying possible quantum computer materials.

4.6. Free Iron and Ageing

In this project I could show that it is possible to measure the increase of free iron in *C. elegans* upon feeding the worms with FAC using EPR spectroscopy. While this suggests that the in organism free iron pool that is available for the Fenton reaction can be altered, it did not show an effect on life span, even though there was an effect on oxidative damage.

In summary the results of the study challenge the established theory of ageing being influenced by oxidative damage of ROS produced by the Fenton reaction.

4.7. Overall Remarks

In view of the thesis title the presented projects mostly consist of a combination of computational and experimental investigation. My research on HDAC8 is an example of how MD simulations can be used to inspire experimental investigations. I did this by the means of structural and dynamics investigations, i.e. my NMR investigations in Section 2.4, or biochemical investigations, i.e. my development and use of a novel NMR based deacetylation assay described in Sections 2.2 and 2.3.

Hence, the main project as a whole is an example of how to combine simulation and experiment successfully to investigate the function of a metalloprotein. The very same goes for the NBO approach described in Section 3.1, where I used

in restrained MD to improve the structural sampling of MD simulation based on NMR measurements. This is a direct use of experimental data to improve the power of computation and therefore possibly improve future investigations of other metalloproteins with a paramagnetic metal.

The material science themed study of Si:Bi is yet another example of how experimental and computational science. Here, an initial experimental characterisation was used as a starting point for CCE calculations to yield a novel hypothesis for an optimal working point of the studied material.

Therefore, the thesis successfully shows what is encapsulated in the title, it shows the power of pairing magnetic resonance spectroscopy methods with computational methods for the investigations of metalloproteins and use in other areas of research.

5. Methods

This chapter provides details and parameters of the techniques used, including protocols for sample preparation and the setup of the simulations conducted. Furthermore, some software that has been written for the work presented and work in the lab are briefly described.

5.1. HDAC8 Sample Preparation

The vector used for protein expression consists of a His tag, NusA, TEV cleavage site and HDAC8, codon optimised. The vector was purchased from GenScript (GenScript USA Inc., 860 Centennial Ave., Piscataway, NJ 08854, USA). Protein concentrations were determined by measuring the A_{280} (optical absorbance at 280 nm on a Biomate 3 from Thermo Scientific with Fischer micro cuvettes with a path length of 1 cm. The extinction coefficient η_{280} for HDAC8 was calculated to be 52120 by using the ProtParam tool at http://expasy.org/. To get single point mutants of HDAC8 I used the QuikChange Site-Directed Mutagenesis Kit as described in products protocol except the DPN1 digest, which was prolonged to last $>$ 3 h. Mutations were made with primer in the pET-29b+ plasmid. Primers were designed using PrimerX (primerx.org), which automatically optimises the primer sequence for recombinant protein expression in *E. coli*. The used primers are listed below. Lys33Glu:

1. F:GTATGTGCGATTCCCTGGCGGCCATCCCGAAACGTGCCTCAATG
2. R:CATTGAGGCACGTTTCGGGATGGCCGCCAGGGAATCGCACATAC

Asp87-89Arg:

1. F:GCAAAAAGTGAGCCAGGAAGGTCGCCGCCGTCATCCGGATTC
2. R:GAATCCGGATGACGGCGGCGACCTTCCTGGCTCACTTTTTGC

The Lys33Glu/Asp87-89Arg mutant was made by using the Lys33Glu primer on the Asp87-89Arg template.

5.1.1. Protein Purification

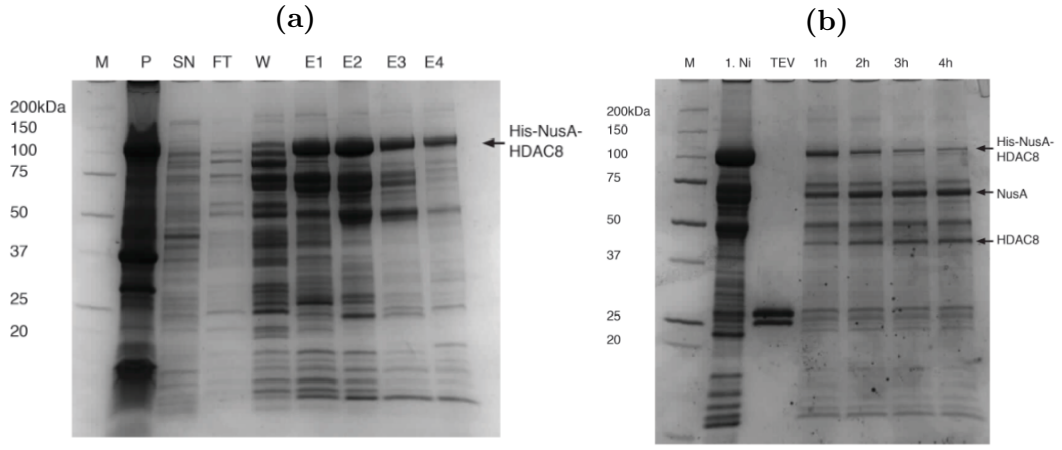


Figure 5.1.: (a) SDS-PAGE of the first Ni-NTA column: Marker (M), Pallet (P), Super Natant (SN), Wash (W), Flow Through (FT) and Elution fractions (E1-4). (b) SDS-PAGE of TEV cleavage: Marker (M), Protein before cleavage after first NI-NTA column (1. Ni), TEV Protease (TEV) and the cleavage product at different time steps (1 h etc.).

BL21(DE3) pET-29b+ His-NusA-TEV-HDAC8 cells were grown in LB media (or M9 media for isotope labelling) at 37 °C and induced by the addition of isopropyl-β-D-thiogalactoside (1 mM) and $ZnSO_4$ (200 μM) when A_{600} (absorbance at 600 nm) was between 0.6 to 1. Cells grown in LB were harvested 3 h after induction. For isotope labelled proteins the temperature was decreased to 23 °C prior to induction and cells were harvested after approximately 16 h. For $C_{\delta 1}$ methyl labelled samples α-keto-buteric acid was used as a precursor with the desired isotopomer methyl group (purchased from Sigma).[85] If the

precursor was not readily deuterated on all positions except the methyl group I used a H-D exchange protocol similar to the one in Ref. 83 to achieve deuteration. Prior to adding the precursor to the bacterial culture deuteration was assessed using ^{1}H NMR. For a 2 L bacterial culture 120 mg of the precursor were dissolved in 400 mL of D_2O and potassium and sodium phosphate buffers were added to match the concentrations in the M9 medium. pH* (i.e. uncorrected pH) was adjusted to 10.2 and the mixture was set slowly shaking at 40 °C o/n. pH* was adjusted to match the pH of the M9 medium and added 20 min before induction.

Cells were harvested by centrifugation (4 °C, 15 min, 4 krpm) on a Beckman and Coulter standing centrifuge using a JLA rotor. Residual cell pallets were resuspended in lysis buffer (50 mM Tris, pH 8.0, 500 mM KCl, 5 mM imidazole, 3 mM $MgCl_2$, 5% glycerol, 5 mM beta-mercaptoethanol, 0.25% IGEPAL) with complete protease inhibitor, lysozyme and DNAse.

Cells were lysed using sonication and centrifuged (4 °C, 50 min, 18 krpm) on a Sorvall standing centrifuge using a SS34 rotor. Supernatant was sterile filtered (0.2 μm) and applied to a Ni-NTA column (GE Healthcare). The column was washed with washing buffer (50 mM Tris, pH 8.0, 500 mM KCl, 5 mM imidazole, 3 mM $MgCl_2$, 5% glycerol, 5 mM beta-mercaptoethanol) and eluted with an increased imidazole concentration of 200 mM. Respective elution fractions (example shown in Figure 5.1) were pooled, diluted and dialysed o/n at 4 °C against 2 L of dialysis buffer 1 (50 mM Tris, pH 8.0, 150 mM KCl, 5 mM TCEP, 5% glycerol). TEV-protease was added (0.01 mg/mg protein) and incubated for 4 h at room temperature (Figure 5.1 (b)). When cleavage was not sufficient more protease was added and incubation time was prolonged (e.g. o/n at 4 °C).

The TEV-protease comprises a His-tag, which lets one remove the protease by applying the protein TEV mixture to a second Ni-NTA after sufficient TEV cleavage. Collecting the flow trough of the second Ni-NTA column yields the desired HDAC8. To increase purity of the protein the sample is applied to a

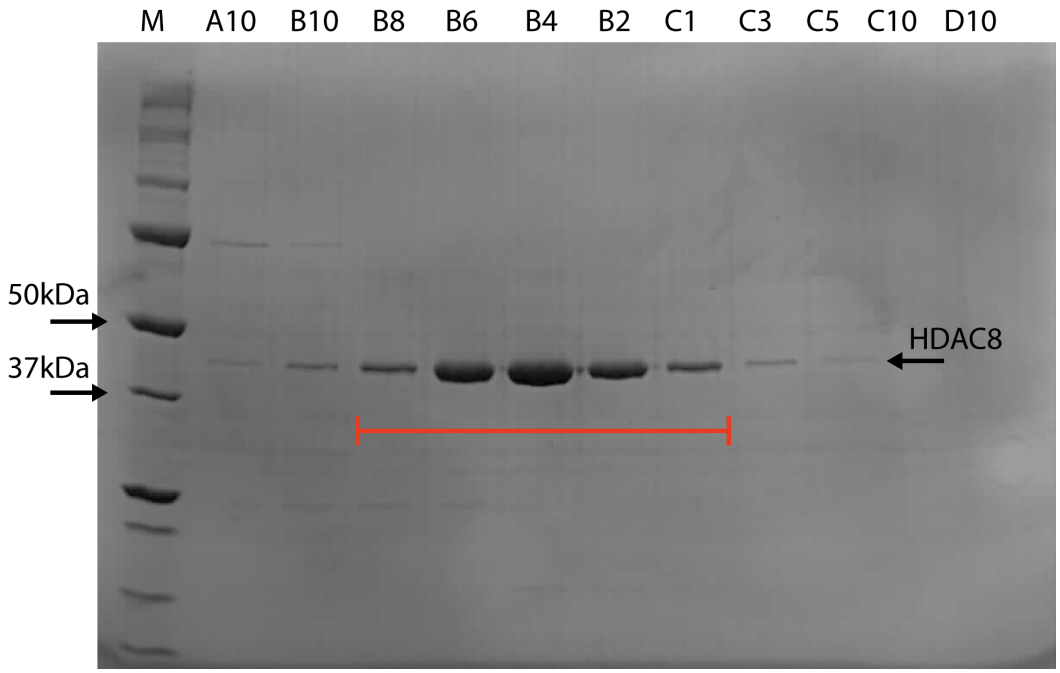


Figure 5.2.: SDS-page of HDAC8 after gel filtration on a Superdex S75 column. The wells are labelled on the top are the Marker (M) and the well from collection. Relevant marker bands are highlighted and labelled, the same ladder was used as in Figure 5.1. The red line marks the pooling to yield a purity >85% as judged by the presented SDS-page.

gel-filtration column (Superdex S75). A SDS-page after the gel-filtration can be seen in Figure 5.2.

5.2. HDAC8 Activity Assays

Catalytic activity of the HDAC8 was tested using the commercially available Fluor de Lys HDAC8 test kit (BIOMOL) or an analogue (MAL) produced by Chris Matthews in the group of Prof. Charles Marson (UCL Chemistry).[150] The Fluor de Lys HDAC8 assay uses a peptide (4 amino acids) substrate structurally based on acetylated p53 connected to a fluorophore (excitation 340 nm). In contrast, the MAL assays uses Boc-Lys(Ac)-7-amino-4-methylcoumarin, i.e. only the acetylated residue is present between a Boc protection group and the

fluorophore. The HDAC inhibitor Trichostatin A (TSA) was purchased from Enzo Life Sciences and porcine pancreatic trypsin (type IX-S) from Sigma.

The procedures is the same for both assays. Adding the substrate to the enzyme the reaction takes place for a certain amount of time and is subsequently quenched with the TSA inhibitor. At the same time trypsin is used to digest the deacetylated compound, which leads to a change in the emission wavelength from 380 to 450 nm. The amount of deacetylated compound is detected by a fluorescence plate reader and therefore gives a measure of the activity or reaction rate.

In detail, aliquots were prepared for HDAC8 wt and each mutant to yield a $0.4, 1$, and $2\,\mu$M final concentration in the total reaction volume of $60\,\mu$L in assay buffer (50 mM Tris at pH 8.0, 137 mM NaCl, 2.7 mM KCl, 1 mM $MgCl_2$, 1 mg/mL BSA). MAL substrate solution of 50 mM in DMSO was diluted in assay buffer and added to the enzyme solution to yield a final concentration of $250\,\mu$M in the reaction volume.

The HDAC8:MAL solution was incubated at 25 °C for 30 or 60 min, after which $50\,\mu$L of the reaction solution was pipetted on a 96-well white NBS microplate, where the wells had been preloaded with $50\,\mu$L developer solution (10 mg/mL trypsin and $4\,\mu$M TSA in assay buffer). The microplate was left for 30 min at ambient temperature before the fluorescence was measured on a BMG FLUOstar Optima plate reader with excitation at 380 nm and emission at 460 nm. Activities for different reaction times and concentrations were compared to the wt enzyme and averaged as summarised in Figure 2.7 (b). Errors were estimated as the root-mean-square deviation of the activity measured in the six assays for each mutant (two time points and three concentrations).

The k_{cat}/K_m for the wt enzyme was determined using substrate concentrations of 25, 50, 100, 200, 400, and $600\,\mu$M and an enzyme concentration of 400 nM. Reactions were incubated for 15, 30, and 45 min (Figure 2.8). The error given represents the 2σ interval for the χ^2 fitting.

As a complementary tool to assess the activity of HDAC8 I used the assay described in Section 2.3. Here, the Michaelis-Menten ODEs are fitted directly from kinetic data. Hence, a brief derivation of the ODEs are given.

5.2.1. Michaelis-Menten Kinetics

Leonor Michaelis and Maud Menten proposed one of the most useful models for enzyme kinetics.[209,210] In this model the enzymatic reaction is described by relating the concentration $[S]$ of the substrate S with the reaction rate v, which is the change of the product concentration $[P]$ over time. This can also be expressed as:

$$v = \frac{d[P]}{dt} = \frac{V_{max}[S]}{K_m + [S]}, \tag{5.1}$$

where V_{max} is the maximum rate of the reaction under saturating conditions. K_m is the so-called Michaelis constant, which represents the substrate concentration at which the reaction rate v equals half of the maximum achievable rate V_{max}. The basic idea of enzyme kinetics in this model involves the enzyme E binding to the substrate S to form a complex ES. This complex can then be converted into the enzyme E and the product P.

Formation of the complex ES and dissociation of enzyme and substrate are reversible and are both represented using a forward k_f and reverse k_r rate constant. Conversion of the complex into the enzyme and product is represented by the catalytic rate constant k_c, whereby this conversion is assumed to be irreversible. Therefore the model can be written as a set of processes:

$$E + S \underset{k_r}{\overset{k_f}{\rightleftharpoons}} ES \xrightarrow{k_c} E + P \tag{5.2}$$

Using the law of mass action one can derive a set of four ordinary differential equations describing the change of the system over time t:

$$\begin{aligned}
\frac{d[E]}{dt} &= -k_f[E][S] + k_r[ES] + k_c[ES] && (5.3)\\
\frac{d[S]}{dt} &= -k_f[E][S] + k_r[ES] && (5.4)\\
\frac{d[ES]}{dt} &= k_f[E][S] - k_r[ES] - k_c[ES] && (5.5)\\
\frac{d[P]}{dt} &= k_c[ES] && (5.6)
\end{aligned}$$

As the enzyme acts as a catalyst, it is clear that the total concentration of enzyme $[E]_0$, the sum of the free enzyme and enzyme in complex with the substrate, is constant. This can also be seen from formulas 5.3 and 5.5, as:

$$\frac{d[E]}{dt} = -\frac{d[ES]}{dt}, \tag{5.7}$$

i.e. $[E]_0$ is conserved.

5.2.2. Equilibrium Approximation

The reaction rate can be derived for this model when assuming a steady-state of association and dissociation of the complex ES, i.e. $k_f[E][S] = k_r[ES]$. Applying this relation to the conservation law in formula 5.7 one can derive the complex concentration $[ES]$:

$$\begin{aligned}
k_r[ES] &= k_f\left([E]_0 - [ES]\right)[S] && (5.8)\\
\frac{k_r[ES]}{k_f[S]} &= [E]_0 - [ES] && (5.9)\\
[ES]\left(\frac{k_r}{k_f[S]} + 1\right) &= [E]_0 && (5.10)\\
[ES] &= \frac{[E]_0}{\left(\frac{k_r + k_f[S]}{k_f[S]}\right)} && (5.11)\\
[ES] &= \frac{[S][E]_0}{\frac{k_r}{k_f} + [S]} = \frac{[S][E]_0}{K_d + [S]}, && (5.12)
\end{aligned}$$

where $K_d = \frac{k_r}{k_f}$ is the dissociation constant for the complex. Under saturating conditions every enzyme will form a complex with a substrate and hence the maximum rate V_{max} is the product of the catalytic rate constant k_c and the total enzyme concentration $[E]_0$. Combined with equation 5.6 the velocity is:

$$v = \frac{d[P]}{dt} = \frac{V_{max}[S]}{K_d + [S]} \tag{5.13}$$

It is clear, that this approximation only holds as long as the equilibration of enzyme and substrate is much faster than the product formation, i.e.:

$$\frac{k_c}{k_r} \ll 1 \tag{5.14}$$

has to be valid.

5.2.3. Limitations

Apart from the assumptions made in the approximation the model itself is of course a simplification. The catalytic step is not irreversible in general and therefore a more precise picture would be:

$$E + S \underset{k_{r1}}{\overset{k_{f1}}{\rightleftharpoons}} ES \underset{k_{r2}}{\overset{k_{f2}}{\rightleftharpoons}} E + P \tag{5.15}$$

One usually neglects k_{r2}, which is the rate constant of product and enzyme converting to substrate and enzyme. k_{r2} depends on the free energy of the reaction, which means that if the energy released during the catalytic reaction is large ($\Delta G \ll 0$) the k_{r2} will be negligible. k_{r2} is also negligible in the case where the substrate concentration is much higher than the product concentration, which is true for most biochemical *in vitro* assays like the ones presented in Section 2.2.

Moreover, another limitation is the use of the law of mass action, which comes with the assumption of three-dimensional diffusion of the interacting masses

(here enzyme and substrate). This is, especially *in vivo*, not generally the case as particles in a crowded and compartmented environment like a cell are not diffusing freely in three dimensions, the dimensionality would be a fractal, which one would have to account for. However, for the *in vitro* assay one works with purified protein and substrate, so that $[E] \ll [S]$ is always true. Additionally the free energy difference for the reaction is very favourable with a proposed $\Delta G = -7.1 \pm 0.6\,\mathrm{kcal\,mol^{-1}}$.[133]

5.2.4. Determination of $\mathbf{V_{max}}$ and $\mathbf{K_m}$

To measure the kinetic rate constants one needs to measure the reaction rate via e.g. a fluorescence assay described earlier (Section 5.2). Typically, several time points are used per concentration point to eliminate bias, which might arise from bad mixing or quenching. Thus, one uses the gradient of the fluorescence signal corrected by background fluorescence as the rate v. V_{max} and K_m can than be calculated using non-linear regression, i.e. fitting both to formula 5.1 (Figure 2.8).

The MAL assay does not allow the measurement of all rates and only lets one extract $\frac{k_{cat}}{K_m}$ as the substrates solubility is smaller than the K_m. A similar approach can be used for the NMR activity assay and simply solve the Michaelis-Menten ODEs and obtain $\frac{k_{cat}}{K_m}$ (Section 2.3) without the need of any linearisation, as e.g. a Linweaver-Burk plot does.

5.3. NMR Methods

1D ^{1}H NMR spectra of HDAC8 wt and mutants presented in Section 2.2 (Figure A.6) and for the NMR deacetylation assay (Section 2.3) have been recorded on a Bruker (Karlsruhe, Germany) Avance 500. This spectrometer operates at a proton Larmor frequency of 500 MHz and is equipped with a room temperature probe and a temperature controller.

All spectra were recorded at 25 °C using excitation sculpting for water suppression.[211] Pulses for all experiments were given at the highest available power. Hence, the approximate frequency for 90° pulses was 30 kHz.

5.3.1. $^{13}CH_2D$ NMR Parameters

$^{13}CH_2D$ experiments have been recorded using a Varian (Agilent Technologies, Santa Clara, CA, USA) Inova 600 operating at a proton Larmor frequency of 600 MHz. Two identical spectrometers have been used, one with and the other one without a cryoprobe equipped. Both spectrometers have temperature controller equipped and all spectra were recorded at 25 °C. Pulses for all experiments were given at the highest available power. Hence, the approximate frequency for pulses was: $\frac{\pi}{2}(^1\mathrm{H}) = 30\,\mathrm{kHz}$; $\frac{\pi}{2}(^{13}\mathrm{C}) = 17\,\mathrm{kHz}$; $\frac{\pi}{2}(\mathrm{D}) = 1.6\,\mathrm{kHz}$. All samples were prepared in 30 mM KP buffer with 30 mM KCl and 1 mM TCEP at pH* 8.0 in pure D_2O.

For measuring the $R^Q(D_z)$ using the HSQC sequence from Section 2.4.2 T_1 and T_{max} are the relaxation time variables. Phase cycling: $\phi_1 = x, -x$; $\phi_2 = 2(x), 2(-x)$; $\phi_3 = 4(x), 4(-x)$ and $\phi_{rec} = x, -x, -x, x, -x, x, x, -x$. Delays: $\tau_a = 1.7\,\mathrm{ms}$; $\tau_d = 3.85\,\mathrm{ms}$. Gradients in (ms, relative strength): $g_1 = (0.5, 5)$; $g_2(0.3, 3)$; $g_3(1.5, 15)$; $g_4(0.3, -12.5)$; $g_5(0.4, -10)$; $g_6(0.2, 10)$; $g_7(1, 15)$; $g_8(0.3, 2)$; $g_9(0.5, 1.35)$. Decoupling is achieved using WALTZ-16. The HSQC spectrum shown in Figure 2.15 has been recorded with 64 increments in the indirect t_1 dimension with a sweep width of 2250 Hz. Direct dimension has been recorded with 576 points with a sweep width of 9009 Hz. Spectra have been recorded for 7.5 h in total. For the comparison to the TROSY spectra only one T_1 value has been used, i.e. $T_1 = 1\,\mathrm{ms}$ and $T_{max} = 81\,\mathrm{ms}$.

The $^{13}CH_2D$ TROSY sequence has been described in detail elsewhere[82 71] and is hence not repeated here. The relaxation delays used for deuterium relaxation measurements are the following. For $R^Q(D_z)$: $T_1 = 2, 10, 21, 32, 45, 59, 80\,\mathrm{ms}$ and $T_{max} = 82\,\mathrm{ms}$. For $R^Q(D_+)$: $T_2 = 0.05, 0.18, 0.38, 0.59,\ 0.82,\ 1.08,\ 1.36,$

1.7 ms and $T_{max} = 2$ ms. For $R^Q(3D_z - 2)$ the same delays as for $R^Q(D_z)$ have been used. For $R^Q(D_+D_z + D_zD_+)$ the same delays as for $R^Q(D_+)$ have been used. Each deuterium relaxation experiment was acquired for 47 h in total.

Relaxation delays were recorded scrambled, i.e. measured in a non-monotonic fashion, as to avoid artefacts due to shimming errors or protein degradation over the course of the experiment. This way, trends in the acquired data that depend on the elapsed time of the experiment are avoided.

5.4. PEPC

Potential Energy Perturbation Calculation (PEPC)[1] is a pipeline I wrote that has been inspired by conversations with Kresten Lindorff-Larsen and the development of FoldX.[30,212,213] The aim of this tool is to investigate conformational changes of proteins by perturbing the potential energy landscape.

More specifically, when looking at a protein that samples different conformational states that are important for a biological function, PEPC can be used to calculate how a single point mutation would likely change the energy of these states. Thus, predicting whether a specific mutation would make a certain state more or less favourable, and thereby get a 're-weighting' of the conformational states, as shown in Section 2.1 (Figure 2.6).

In very simple terms the idea can be described as follows: Assume that the energy of a certain time dependent protein structure in a MD simulation is $E_{wt}(t)$, the energy of the wt structure. Introducing a mutation to the structure changes the energy to E_{mutant}. The resulting response, i.e. $\Delta E = E_{mut} - E_{wt}$ will then report whether or not this particular structure (state) will be more or less energetically favourable. Hence, $\Delta E(t) = E_{mut}(t) - E_{wt}(t)$ estimates a 're-weighting' of certain structures or states.

In principle PEPC can also be used *vice versa* to define states by simply using the perturbed energy profiles to cluster the conformations. Additionally,

running PEPC on a wt MD trajectory is much faster than re-running the simulation with a single point mutation, while eliminating some ambiguity of mutant versus wild type simulation due to the stochastic nature of MD.

PEPC is written as a set of Python (www.python.org) wrapper scripts that make use of the GROMACS[187,214–216] software and MMTSB[217] scripts to mutate protein structures and calculate its energies. Optionally this can also be done using the FoldX engine, as has been done for Figure 2.6. In this way an estimate for the free energy change is generated rather than for the potential energy.

PEPC comes as a library of python scripts that are also able to invoke docking calculations, which allow docking of a ligand to all different conformations and mutants.

5.4.1. Programme Flow

In principle the work flow has four steps: 1) extracting the structure (frame) 2) mutating the desired residue and 3) calculating the potential energy 4) readout. In the case of a MD trajectory steps 1)-3) are repeated many times, therefore the scripts are by default organised with a certain data structure shown in figure 5.3. Scripts can be used UNIX-style in a terminal window and display all options when using the help flag (-h). By default they also write (or take as input) configuration files for easier consecutive use.

After doing a wt simulation, frames are extracted using *extract_frames.py*, which saves the specified frames in PDB format in a folder. All structures are then mutated using *mutate_frames.py* (using MMTSB[217] or FoldX routines). In the next step the single point potential energy of the (mutated) structure is calculated with *gmx_sp_energy.py*, which creates a topology for each structure and runs a single point MD simulation and writes out the energy terms to a

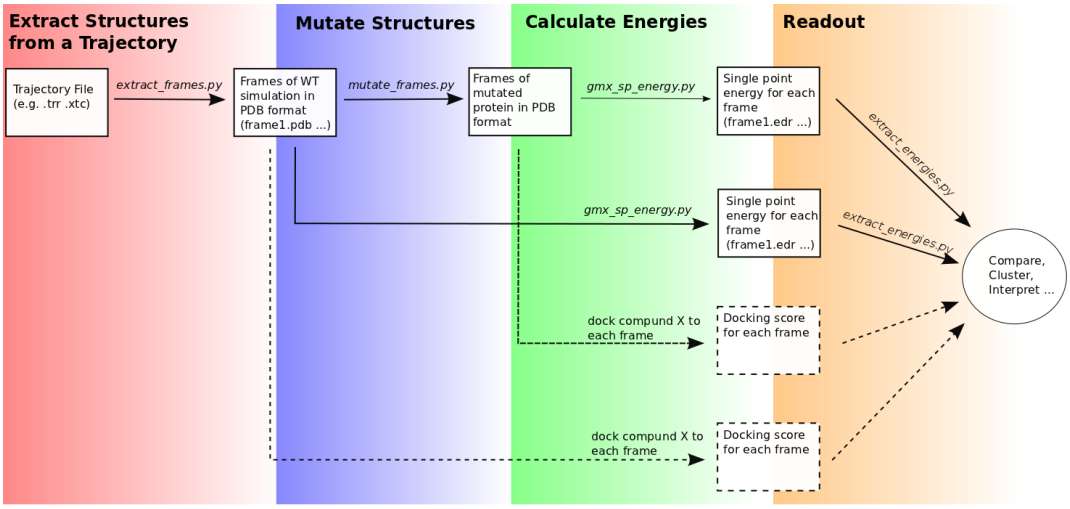


Figure 5.3.: Flow chart of PEPC highlighting the four steps for the calculations: 1) extraction of frames (red) 2) mutating structures (blue) 3) calculating energies (green) and 4) reading out energy terms (orange). Scripts that are used for the individual steps are written in italic letters.

text file. Specified energy terms (e.g. short range electrostatic) are then read out by *extract_sp_energies.py* and saved as a single text file.

It is noteworthy that to get comparable energies one has to calculate the energies for the wild type with this programme as well, as the energy will depend on the MD input parameters. By default, the programme uses infinite cut-offs, implicit solvent and no periodic boundary conditions, as this is generally the only way to remove energy bias arising from explicit water molecule clashes or even a different number of placed water molecules in the simulation box during topology set up. This in turn also means that the absolute energy is not comparable to running simulations in explicit water with periodic boundary conditions.

As a consequence only the difference in energy (the response to the perturbation) can be interpreted. However, this has to be done carefully as PEPC is only designed to give a rough estimate of expected changes.

5.4.2. Considerations and Limitations

The most obvious and severe limitation is that one uses a wt trajectory to calculate the mutant energies. Whereas high energies show that a state is unfavourable, in a 'real' mutant MD simulation the protein might be able to accommodate the mutation by other subtle changes in conformation, which one might never observe in a wild type simulation. A mutation that leads to an unfolded protein is also outside the scope of PEPC. Hence, one has to assume that the mutation is only of local nature and leaves the overall architecture of the protein intact.

While this is not always the case, often enough it is and especially for larger proteins (>200 amino acids) this is likely to be the case for many single point mutations.[212]

5.5. Average Solvent Distance Calculations

Paramagnetic relaxation enhancement (PRE) of nuclear spins by electron spins depends on the distance r between the two spins, i.e. the PRE effect scales with r^{-6}. The phenomenon of PRE can be exploited in many ways, e.g to measure distances in paramagnetically labelled proteins by using a nitroxide or lanthanide tag[218] and therefore as constraints in MD simulations.[219] It can also be used for aiding the assignment of residues by exploiting the fact that each residue's solvent exposure can be slightly different, depending on the positioning of the atom in the protein and the shape of the protein itself. In such experiments a PRE agent is added to the solvent that enhances the relaxation of the nuclear spins in the protein, giving rise to a characteristic relaxation enhancement pattern.

The solvent accessible surface area (SASA) of certain atoms can be easily calculated with standard molecular graphics tools such as PyMOL (Schrödinger LLC) or VMD.[220] However, for PRE the actual average solvent distance is

important and not only the SASA as the protein can have tunnels or cavities etc. that lead to high SASA values for an atom without being accessible to the bulk. The whole solvent has to be taken into account, at least up to a distance after which the PRE effect is negligible.

Here, I describe a tool that easily and quickly calculates the average solvent distance (ASD) for r^{-1} and r^{-6} weights for proteins with solved structure. This I realised in Python with the use of the MDAnalysis[221] toolkit (Appendix A.8). A brief work flow of the programme will be given with an illustration of the computational concepts in Figure 5.4.

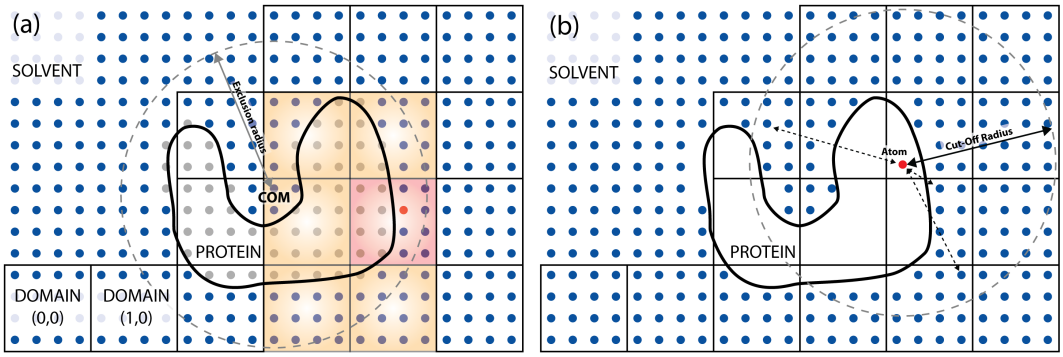


Figure 5.4.: 2D illustration of ASD (average solvent distance) calculations. (a) Schematic representation of how the protein imprint (protein shown as a schematic 2D cross section in the middle) is calculated. Blue dots represent solvent grid points and black rectangles represent the domains that are tagged with (x, y) indices . The protein COM is highlighted as well as the exclusion radius. Domain-decomposition is exemplified by the red solvent point, the closest protein loin is first searched in the same(red shaded domain) and if below the minimum distance in the neighbouring domains (orange shaded) within the exclusion radius. Grey dots signify solvent points that are removed for the imprint. (b) Most of the representations is as in (a). However, here the red dot shows a protein atom for which the ASD is calculated within a cut-off radius (grey dashed circle). Dashed arrows illustrate the summation over all these solvent point to protein atom distances.

5.5.1. Modelling the Solvent

The first problem is modelling the solvent, which is achieved by a tuneable linear solvent grid (blue dots in Figure 5.4). In the next step an imprint of the protein is calculated, i.e. the solvent grid points that are within the protein (within a certain cut-off) are removed from the grid (tray dots in Figure 5.4 (a)).

In fact, not all solvent point to protein atom distances are calculated, the number of computations being drastically reduced by the combination of two approaches, domain decomposition and an exclusion radius. First the solvent grid is divided into cubic domains (black rectangles in Figure 5.4 (a)), which is also done for the protein and the same domains. The centre of mass (COM) of the protein is calculated and the maximum distance to a heavy atom of the protein is calculated and taken as exclusion radius. Then the distance of each solvent domain geometric centre (GC) to the protein COM is calculated and excluded from imprint calculations if this distance (plus a buffer distance - domain GC to domain corner distance) is larger than the exclusion radius.

After this, the minimum distance of each solvent point from each domain within the exclusion radius to the protein atoms has to be found in order get the protein imprint. This is achieved by first calculating the distance of each solvent point in a domain (e.g. the red dot in the red shaded rectangle in Figure 5.4 (a)) to each protein atom within the same protein domain (the red rectangle in Figure 5.4 (a)). If the minimum distance is above the set minimum distance (atom centre to atom centre for protein atom to PRE agent) within the domain the neighbouring protein domains that are also within the exclusion radius (orange shaded domains in Figure 5.4 (a)) are searched.

Depending on the grid size, protein size and grid spacing this drastically speeds up the calculation compared to a brute force distance calculation, i.e. all solvent points versus all protein atom points.

5.5.2. Calculating the ASD

After the imprint has been calculated (blue dots in Figure 5.4 (b)) the ASD has to be measured. This is done with domain-decomposition as well, i.e. only solvent points in domains whose GCs are within the cut-off radius (plus buffer distance) are concerned. The distances for each nucleus to solvent point (dashed arrows in Figure 5.4 (b)) are calculated, weighted (r^{-1} and r^{-6}) and summed up for each nucleus of interest, i.e. let x be the position of the atom (red dot in Figure 5.4 (b)) and s_n the n-th solvent point position within the cut-off radius, the solvent weight w for r^{-6} weighting is calculated as:

$$w = \sum_{n}^{N} (dist(x, s_n))^{-6}, \tag{5.16}$$

where N is the total number of solvent points within the cut-off radius.

Using these simple approaches a PRE pattern for any nuclei selection in a given protein structure can be calculated. For a large grid spacing (>1 Å) the results are heavily discretised and error prone, due to the weighting factors. However, using a fine enough grid the solutions of the programme converge (<0.25 Å). The calculations are simple with an automatic grid setup and just need a supplied PDB file and selection of nuclei.

5.6. Molecular Dynamics Simulations

The first set of unbiased MD simulations presented in Section 2.1 were performed using GROMACS 4.5.3[187,214–216] with double precision, the AMBER99SB-ILDN[56] force field and TIP3P water. Five missing residues in 1T69 were initially modelled using Modeller 9.09.[222] The solvated systems consisted of approximately 57000 atoms. 0.9 nm cut-offs were used for electrostatic and van der Waals interactions and the Particle Mesh Ewald (PME) method was applied for long-range electrostatics.[34]

I used an extensive equilibration protocol, where each system was simulated for 5 ns in an NVT ensemble followed by 5 ns in an NpT ensemble during which positional restraints of 1000 kJ mol^{-1} nm^{-2} were applied on heavy atoms of the protein and increased the temperature. All production runs were performed in an NpT ensemble using a Nosé-Hover thermostat at 300 K, an isotropic Parrinello-Rahman pressure coupling and periodic boundary conditions. The neighbour list was updated every 5 steps using 2 fs integration time steps while keeping bonds involving hydrogens and heavy atoms constrained using the P-LINCS algorithm.[187]

The topology for SAHA was generated using Amber-Tools 1.5 with the RESP methodology after geometry optimisation at the B3LYP 6-31G* level using ORCA 2.8.0.[223] Protonation states of histidines were calculated using the H++ server[224] and the protonation states of histidines 142 and 143 close to the catalytic site were taken from Wu et al.[133] as protonated on N_δ.

Analysis of backbone angles was done in VMD 1.9.[220] Data was analysed using GROMACS, VMD 1.9 and MATLAB (MathWorks). Sequence similarities and structure similarity Z-score of crystal structures were obtained from the Dali server.[225] Secondary structure analysis was done using STRIDE.[226] Protein structures were rendered in PyMOL 1.4.1 (Schrödinger LLC.).

The Anton[20,48] production simulations were run in the NpT ensemble with the same force field and water model as the GROMACS simulations. 1.1 nm cut-offs were used for short range interactions and a 64x64x64 PME mesh was used for long range electrostatic interactions. A Berendsen thermostat and barostat[33] were applied to maintain 300 K and 1 bar as the simulation temperature and pressure, respectively.

Postface

I thoroughly enjoyed my time as a Ph. D. student in the Hansen Lab and at UCL in general. I am grateful for the financial support of the Wellcome Trust, the BBSRC and the NIH who all contributed to my work. There are many great people who have either contributed to the presented work or supported me in one way or another and I would like all of them to know that I am very grateful for that.

Flemming Hansen, it was truly great working with you. There are so many aspects I could mention that I am bound to miss several of them in any list I would attempt to compile. Hence, I just mention that I enjoyed 'fluffing' about in the lab with you, working on biochemistry/ quantum chemistry/ spin physics and statistical mechanics. You are awesomely smart! As a small token of my gratitude I hereby give you the copyrights for the Hansen Lab logo I created.

Chris Kay, I will always be grateful for your continuous support, without which I would have not been able to embark on this Ph. D. programme. You supported me since I arrived at UCL before my time in this programme. I value you as a scientist as well as as a friend.

Peter Coveney, I would like to thank you for your support and guidance in many aspects of my work. I admire your vision for research and it was great to a part of that vision.

Nico Werbeck, you are the best colleague I could wish for and an incredible friend. It was great working with you, but it was even better to get inspired by you during work and during our coffee breaks. While we are working on our

'Note on Biology' I thought we might write a book titled 'A History of the Lab in 100 Discussions' in the mean time. In fact, we could 'just make it bigger' and make it 1000 Discussions. On second thought, that might be a bit 'dense'.

Andrea Sauerwein, thank you for being a great comrade in the lab and always joining for coffee - I am refusing at this point to admit that you actually drink tea.

John Kirkpatrick, thank you for your help with the NMR spectrometers and all the patient explanations over the years. I would have certainly been lost several times if it wasn't for you. By the way, you are amazing at explaining NMR (especially constant time).

Simon Greenwood and Chris Matthews, even though you guys haven't been around for a bit. Thank you for lifting my spirits on many occasions and constant delivery of real life 'peacock' stories (at least from Simon). I would also like to thank you guys for supplying me with HDAC8 inhibitors and MAL substrate.

I would also like to thank all present and past members of the Hansen Lab I have worked with and all the colleagues I have met on the 6th floor in the Darwin Building (especially since there is 'cake monday').

Svenja, thank you for being there for me throughout all the time over such a long distance. I am very much looking forward to reduce this distance very soon. Kuss.

Christoph, I hope you still 'take it from' and thereby reach the 'top'.

Last but not least I would like to thank the rest of my family and friends for their support.

Bibliography

[1] Kunze, M. B. A.; Wright, D. W.; Werbeck, N. D.; Kirkpatrick, J.; Coveney, P. V.; Hansen, D. F. Loop interactions and dynamics tune the enzymatic activity of the human histone deacetylase 8. *Journal of the American Chemical Society* **2013**, *135*, 17862–8.

[2] Hansen, D. F.; Westler, W. M.; Kunze, M. B. A.; Markley, J. L.; Weinhold, F.; Led, J. J. Accurate structure and dynamics of the metal-site of paramagnetic metalloproteins from NMR parameters using natural bond orbitals. *Journal of the American Chemical Society* **2012**, *134*, 4670–82.

[3] Valentini, S.; Cabreiro, F.; Ackerman, D.; Alam, M. M.; Kunze, M. B. A.; Kay, C. W. M.; Gems, D. Manipulation of in vivo iron levels can alter resistance to oxidative stress without affecting ageing in the nematode C. elegans. *Mechanisms of Ageing and Development* **2012**, *133*, 282–90.

[4] Phan, G. et al. Crystal structure of the FimD usher bound to its cognate FimC-FimH substrate. *Nature* **2011**, *474*, 49–53.

[5] Balian, S. J.; Kunze, M. B. A.; Mohammady, M. H.; Morley, G. W.; Witzel, W. M.; Kay, C. W. M.; Monteiro, T. S. Measuring central-spin interaction with a spin-bath by pulsed ENDOR: Towards suppression of spin diffusion decoherence. *Physical Review B* **2012**, *86*, 104428.

[6] Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins. *Nature* **2007**, *450*, 964–72.

[7] Hansen, D. F.; Vallurupalli, P.; Kay, L. E. Using relaxation dispersion

NMR spectroscopy to determine structures of excited, invisible protein states. *Journal of Biomolecular NMR* **2008**, *41*, 113–20.

[8] Hansen, D. F.; Vallurupalli, P.; Lundström, P.; Neudecker, P.; Kay, L. E. Probing chemical shifts of invisible states of proteins with relaxation dispersion NMR spectroscopy: how well can we do? *Journal of the American Chemical Society* **2008**, *130*, 2667–75.

[9] Boehr, D. D.; Dyson, H. J.; Wright, P. E. An NMR perspective on enzyme dynamics. *Chemical Reviews* **2006**, *106*, 3055–79.

[10] Wan, S.; Coveney, P. V. Molecular dynamics simulation reveals structural and thermodynamic features of kinase activation by cancer mutations within the epidermal growth factor receptor. *Journal of Computational Chemistry* **2011**, *32*, 2843–52.

[11] Bhabha, G.; Lee, J.; Ekiert, D. C.; Gam, J.; Wilson, I. a.; Dyson, H. J.; Benkovic, S. J.; Wright, P. E. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* **2011**, *332*, 234–8.

[12] Godoy-Ruiz, R.; Guo, C.; Tugarinov, V. Alanine methyl groups as NMR probes of molecular structure and dynamics in high-molecular-weight proteins. *Journal of the American Chemical Society* **2010**, *132*, 18340–50.

[13] Corbi-Verge, C.; Marinelli, F.; Zafra-Ruano, A.; Ruiz-Sanz, J.; Luque, I.; Faraldo-Gómez, J. D. Two-state dynamics of the SH3-SH2 tandem of Abl kinase and the allosteric role of the N-cap. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, E3372–80.

[14] Boehr, D. D.; McElheny, D.; Dyson, H. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **2006**, *313*, 1638–42.

[15] Adamczyk, A. J.; Cao, J.; Kamerlin, S. C. L.; Warshel, A. Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 14115–20.

[16] Rimratchada, S.; McLeish, T. C. B.; Radford, S. E.; Paci, E. The role of high-dimensional diffusive search, stabilization, and frustration in protein folding. *Biophysical Journal* **2014**, *106*, 1729–40.

[17] Bouvignies, G.; Vallurupalli, P.; Hansen, D. F.; Correia, B. E.; Lange, O. F.; Bah, A.; Vernon, R. M.; Dahlquist, F. W.; Baker, D.; Kay, L. E. Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **2011**, *477*, 111–4.

[18] Palmer, A. G.; Massi, F. Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chemical Reviews* **2006**, *106*, 1700–19.

[19] Haider, S.; Joseph, C.; Neidle, S.; Fierke, C. A. On the function of the internal cavity of histone deacetylase protein 8: R37 is a crucial residue for catalysis. *Bioorganic & Medicinal Chemistry Letters* **2011**, *21*, 2129–32.

[20] Shaw, D. E. et al. Millisecond-scale molecular dynamics simulations on Anton. Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis - SC '09. New York, New York, USA, 2009; p 1.

[21] Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334*, 517–20.

[22] Freddolino, P. L.; Arkhipov, A. S.; Larson, S. B.; McPherson, A.; Schul-

ten, K. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure* **2006**, *14*, 437–49.

[23] Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B* **2007**, *111*, 7812–24.

[24] Nyden, M. R.; Forney, G. P.; Brown, J. E. Molecular modeling of polymer flammability: application to the design of flame-resistant polyethylene. *Macromolecules* **1992**, *25*, 1658–1666.

[25] Car, R.; Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Physical Review Letters* **1985**, *55*, 2471–4.

[26] Morse, P. M. Diatomic molecules according to the wave mechanics. II. Vibrational levels. *Physical Review* **1929**, *34*, 57–64.

[27] Buckingham, R. A. The classical equation of state of gaseous helium, neon and argon. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1938**, *168*, 264–83.

[28] Jones, J. E. On the determination of molecular fields. II. From the equation of state of a gas. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **1924**, *106*, 463–77.

[29] Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic validation of protein force fields against experimental data. *PloS One* **2012**, *7*, 1–6.

[30] Piana, S.; Sarkar, K.; Lindorff-Larsen, K.; Guo, M.; Gruebele, M.; Shaw, D. E. Computational design and experimental testing of the fastest-folding β-sheet protein. *Journal of Molecular Biology* **2011**, *405*, 43–8.

[31] Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How robust are protein folding simulations with respect to force field parameterization? *Biophysical Journal* **2011**, *100*, L47–9.

[32] Verlet, L. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review* **1967**, *159*, 98–103.

[33] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984**, *81*, 3684–90.

[34] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995**, *103*, 8577–93.

[35] Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981**, *52*, 7182.

[36] Car, R.; Parrinello, M. Structural, dymanical, and electronic properties of amorphous silicon: An ab initio molecular-dynamics Study. *Physical Review Letters* **1988**, *60*, 204–7.

[37] Iannuzzi, M.; Laio, A.; Parrinello, M. Efficient exploration of reactive potential energy surfaces using Car-Parrinello molecular dynamics. *Physical Review Letters* **2003**, *90*, 238302.

[38] Gervasio, F. L.; Carloni, P.; Parrinello, M. Electronic structure of wet DNA. *Physical Review Letters* **2002**, *89*, 108102.

[39] Piana, S.; Carloni, P.; Parrinello, M. Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease. *Journal of Molecular Biology* **2002**, *319*, 567–83.

[40] Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible docking in solution using metadynamics. *Journal of the American Chemical Society* **2005**, *127*, 2600–7.

[41] Bonomi, M.; Parrinello, M. Enhanced sampling in the well-tempered ensemble. *Physical Review Letters* **2010**, *104*, 190601.

[42] Bussi, G.; Laio, A.; Parrinello, M. Equilibrium free energies from nonequilibrium metadynamics. *Physical Review Letters* **2006**, *96*, 090601.

[43] Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *The Journal of Physical Chemistry B* **2005**, *110*, 3533–9.

[44] Pierce, L. C. T.; Salomon-Ferrer, R.; Augusto F de Oliveira, C.; McCammon, J. A.; Walker, R. C. Routine access to millisecond time scale events with accelerated molecular dynamics. *Journal of Chemical Theory and Computation* **2012**, *8*, 2997–3002.

[45] Buch, I.; Giorgino, T.; De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 10184–9.

[46] Ohmura, I.; Morimoto, G.; Ohno, Y.; Hasegawa, A.; Taiji, M. MDGRAPE-4: a special-purpose computer system for molecular dynamics simulations. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* **2014**, *372*, in press.

[47] Shaw, D. E. et al. Anton, a special-purpose machine for molecular dynamics simulation. *Communications of the ACM* **2008**, *51*, 91–97.

[48] Shaw, D. E.; Deneroff, M. M.; Dror, R. O. Anton, a special-purpose machine for molecular dynamics simulation. *ACM* **2007**, 1–12.

[49] Brush, S. G. History of the Lenz-Ising model. *Reviews of Modern Physics* **1967**, *39*, 883–893.

[50] Pronk, S.; Lindahl, E.; Larsson, P.; Pouya, I.; Bowman, G. R.; Haque, I. S.; Beauchamp, K.; Hess, B.; Pande, V. S.; Kasson, P. M. Copernicus: A new paradigm for parallel adaptive molecular dynamics. Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis on - SC '11. New York, New York, USA, 2011; p 1.

[51] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **1983**, *4*, 187–217.

[52] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **1995**, *117*, 5179–97.

[53] Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. The GROMOS biomolecular simulation program package. *The Journal of Physical Chemistry A* **1999**, *103*, 3596–607.

[54] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926.

[55] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of

improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 712–725.

[56] Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–8.

[57] Best, R. B.; Hummer, G. Optimized molecular dynamics force fields applied to the helix-coil Transition of polypeptides. *The Journal of Physical Chemistry B* **2009**, *113*, 9004–15.

[58] Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *The Journal of chemical physics* **2004**, *120*, 9665–78.

[59] Bachmann, S. J.; van Gunsteren, W. F. An improved simple polarisable water model for use in biomolecular simulation. *The Journal of Chemical Physics* **2014**, *141*, 22D515.

[60] Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. The Polarizable Atomic Multipole-based AMOEBA Force Field for Proteins. *Journal of chemical theory and computation* **2013**, *9*, 4046–4063.

[61] Laio, A.; Parrinello, M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 12562–6.

[62] Sutto, L.; Gervasio, F. L. Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 10616–21.

[63] Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. Adaptive biasing force

method for scalar and vector free energy calculations. *The Journal of Chemical Physics* **2008**, *128*, 144120.

[64] Sugita, Y.; Okamoto, Y. Replica exchange molecular dynamics method for protein folding simulation. *Chemical Physics Letters* **1999**, *314*, 141–51.

[65] Bonomi, M.; Barducci, A.; Parrinello, M. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *Journal of Computational Chemistry* **2009**, *30*, 1615–21.

[66] Deighan, M.; Bonomi, M.; Pfaendtner, J. Efficient simulation of explicitly solvated proteins in the well-tempered ensemble. *Journal of Chemical Theory and Computation* **2012**, *2*, 8–11.

[67] Hansen, D. F.; Led, J. J. Determination of the geometric structure of the metal site in a blue copper protein by paramagnetic NMR. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 1738–43.

[68] Hansen, D. F.; Vallurupalli, P.; Kay, L. E. Measurement of methyl group motional parameters of invisible, excited protein states by NMR spectroscopy. *Journal of the American Chemical Society* **2009**, *131*, 12745–54.

[69] Fiaux, J.; Bertelsen, E. B.; Horwich, A. L.; Wüthrich, K. NMR analysis of a 900K GroEL GroES complex. *Nature* **2002**, *418*, 207–11.

[70] Otten, R.; Villali, J.; Kern, D.; Mulder, F. a. a. Probing microsecond time scale dynamics in proteins by methyl (1)H Carr-Purcell-Meiboom-Gill relaxation dispersion NMR measurements. Application to activation of the signaling protein NtrC(r). *Journal of the American Chemical Society* **2010**, *132*, 17004–14.

[71] Tugarinov, V.; Ollerenshaw, J. E.; Kay, L. E. Probing side-chain dynamics in high molecular weight proteins by deuterium NMR spin relaxation: an application to an 82-kDa enzyme. *Journal of the American Chemical Society* **2005**, *127*, 8214–25.

[72] Shiltagh, N.; Kirkpatrick, J.; Cabrita, L. D.; McKinnon, T. A. J.; Thalassinos, K.; Tuddenham, E. G. D.; Hansen, D. F. Solution structure of the major factor VIII binding region on von Willebrand factor. *Blood* **2014**, *123*, 4143–51.

[73] Palmer, A. G. NMR characterization of the dynamics of biomacromolecules. *Chemical Reviews* **2004**, *104*, 3623–40.

[74] Millet, O.; Muhandiram, D. R.; Skrynnikov, N. R.; Kay, L. E. Deuterium spin probes of side-chain dynamics in proteins. 1. Measurement of five relaxation rates per deuteron in 13C-labeled and fractionally 2H-enriched proteins in solution. *Journal of the American Chemical Society* **2002**, *124*, 6439–48.

[75] Tolman, J. R.; Ruan, K. NMR residual dipolar couplings as probes of biomolecular dynamics. *Chemical Reviews* **2006**, *106*, 1720–36.

[76] Guerry, P.; Salmon, L.; Mollica, L.; Ortega Roldan, J.-L.; Markwick, P.; van Nuland, N. A. J.; McCammon, J. A.; Blackledge, M. Mapping the population of protein conformational energy sub-states from NMR dipolar couplings. *Angewandte Chemie International Edition* **2013**, *52*, 3181–5.

[77] Vallurupalli, P.; Hansen, D. F.; Lundström, P.; Kay, L. E. CPMG relaxation dispersion NMR experiments measuring glycine 1H alpha and 13C alpha chemical shifts in the invisible excited states of proteins. *Journal of Biomolecular NMR* **2009**, *45*, 45–55.

[78] Wüthrich, K. *NMR of proteins and nucleic acids*; Wiley, 1986.

[79] S{\o}rensen, O. W.; Eich, G. W.; Levitt, M. H.; Bodenhausen, G.; Ernst, R. R. Product operator formalism for the description of NMR pulse experiments. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1984**, *16*, 163–92.

[80] Prestegard, J. H.; Agard, D. A.; Moremen, K. W.; Lavery, L. A.; Morris, L. C.; Pederson, K. Sparse labeling of proteins: structural characterization from long range constraints. *Journal of Magnetic Resonance* **2014**, *241*, 32–40.

[81] Luginbühl, P. Semi-classical nuclear spin relaxation theory revisited for use with biological macromolecules. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2002**, *40*, 199–247.

[82] Sheppard, D.; Sprangers, R.; Tugarinov, V. Experimental approaches for NMR studies of side-chain dynamics in high-molecular-weight proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2010**, *56*, 1–45.

[83] Tugarinov, V.; Kay, L. E. Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *Journal of the American Chemical Society* **2003**, *125*, 13868–78.

[84] Kay, L. E. NMR studies of protein structure and dynamics - a look backwards and forwards. *Journal of Magnetic Resonance* **2011**, *213*, 492–4.

[85] Ruschak, A. M.; Kay, L. E. Methyl groups as probes of supra-molecular structure, dynamics and function. *Journal of Biomolecular NMR* **2010**, *46*, 75–87.

[86] Lundström, P.; Vallurupalli, P.; Hansen, D. F.; Kay, L. E. Isotope labeling methods for studies of excited protein states by relaxation dispersion NMR spectroscopy. *Nature Protocols* **2009**, *4*, 1641–8.

[87] Tugarinov, V.; Hwang, P. M.; Ollerenshaw, J. E.; Kay, L. E. Cross-correlated relaxation enhanced 1H-13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *Journal of the American Chemical Society* **2003**, *125*, 10420–8.

[88] Ollerenshaw, J. E.; Tugarinov, V.; Kay, L. E. Methyl TROSY: explanation and experimental verification. *Magnetic Resonance in Chemistry* **2003**, *41*, 843–52.

[89] Lipari, G.; Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *Journal of the American Chemical Society* **1982**, *104*, 4546–59.

[90] Lipari, G.; Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *Journal of the American Chemical Society* **1982**, *104*, 4559–70.

[91] Lipari, G.; Szabo, A. Effect of librational motion on fluorescence depolarization and nuclear magnetic resonance relaxation in macromolecules and membranes. *Biophysical Journal* **1980**, *30*, 489–506.

[92] Werbeck, N. D.; Kirkpatrick, J.; Hansen, D. F. Probing arginine side-chains and their dynamics with carbon-detected NMR spectroscopy: Application to the 42 kDa human histone deacetylase 8 at high pH. *Angewandte Chemie International Edition* **2013**, *52*, 3145–7.

[93] Barth, T. K.; Imhof, A. Fast signals and slow marks: the dynamics of histone modifications. *Trends in Biochemical Sciences* **2010**, *35*, 618–26.

[94] Lee, J.-S.; Smith, E.; Shilatifard, A. The language of histone crosstalk. *Cell* **2010**, *142*, 682–5.

[95] Yan, W.; Liu, S.; Xu, E.; Zhang, J.; Zhang, Y.; Chen, X. Histone deacetylase inhibitors suppress mutant p53 transcription via histone deacetylase 8. *Oncogene* **2013**, *32*, 599–609.

[96] Lin, Y.-y.; Kiihl, S.; Suhail, Y.; Liu, S.-Y.; Chou, Y.-h.; Kuang, Z.; Lu, J.-y.; Khor, C. N.; Lin, C.-L.; Bader, J. S.; Irizarry, R.; Boeke, J. D. Functional dissection of lysine deacetylases reveals that HDAC1 and p300 regulate AMPK. *Nature* **2012**, *482*, 251–5.

[97] Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J.-P.; Barsyte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Müller, S.; Pawson, T.; Gingras, A.-C.; Arrowsmith, C. H.; Knapp, S. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **2012**, *149*, 214–31.

[98] Asangani, I. A. et al. Therapeutic targeting of BET bromodomain proteins in castration-resistant prostate cancer. *Nature* **2014**, *510*, 278–82.

[99] Jackson, V.; Shires, A.; Chalkley, R.; Granner, D. K. Studies on highly metabolically active acetylation and phosphorylation of histones. *Journal of Biological Chemistry* **1975**, *250*, 4856–63.

[100] Chestier, A.; Yaniv, M. Rapid turnover of acetyl groups in the four core histones of simian virus 40 minichromosomes. *Proceedings of the National Academy of Sciences* **1979**, *76*, 46–50.

[101] Zee, B. M.; Levin, R. S.; Xu, B.; LeRoy, G.; Wingreen, N. S.; Garcia, B. A. In vivo residue-specific histone methylation dynamics. *Journal of Biological Chemistry* **2010**, *285*, 3341–50.

[102] Ekwall, K. Genome-wide analysis of HDAC function. *Trends in Genetics* **2005**, *21*, 608–15.

[103] Vannini, A.; Volpari, C.; Filocamo, G.; Casavola, E. C.; Brunetti, M.; Renzoni, D.; Chakravarty, P.; Paolini, C.; De Francesco, R.; Gallinari, P.;

Steinkühler, C.; Di Marco, S. Crystal structure of a eukaryotic zinc-dependent histone deacetylase, human HDAC8, complexed with a hydroxamic acid inhibitor. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 15064–9.

[104] Hu, E.; Chen, Z.; Fredrickson, T.; Zhu, Y.; Kirkpatrick, R.; Zhang, G. F.; Johanson, K.; Sung, C. M.; Liu, R.; Winkler, J. Cloning and characterization of a novel human class I histone deacetylase that functions as a transcription repressor. *The Journal of Biological Chemistry* **2000**, *275*, 15254–64.

[105] Marks, P.; Rifkind, R. a.; Richon, V. M.; Breslow, R.; Miller, T.; Kelly, W. K. Histone deacetylases and cancer: causes and therapies. *Nature Reviews. Cancer* **2001**, *1*, 194–202.

[106] Taunton, J.; Hassig, C. A.; Schreiber, S. L. A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* **1996**, *272*, 408–11.

[107] Brownell, J. E.; Zhou, J.; Ranalli, T.; Kobayashi, R.; Edmondson, D. G.; Roth, S. Y.; Allis, C. D. Tetrahymena histone acetyltransferase A : A homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **1996**, *84*, 843–51.

[108] Song, J.; Noh, J. H.; Lee, J. H.; Eun, J. W.; Ahn, Y. M.; Kim, S. Y.; Lee, S. H.; Park, W. S.; Yoo, N. J.; Lee, J. Y.; Nam, S. W. Increased expression of histone deacetylase 2 is found in human gastric cancer. *APMIS : Acta Pathologica, Microbiologica, et Immunologica Scandinavica* **2005**, *113*, 264–8.

[109] Casas-Delucchi, C. S.; Brero, A.; Rahn, H.-P.; Solovei, I.; Wutz, A.; Cremer, T.; Leonhardt, H.; Cardoso, M. C. Histone acetylation controls the

inactive X chromosome replication dynamics. *Nature Communications* **2011**, *2*, 222.

[110] Feng, D.; Liu, T.; Sun, Z.; Bugge, A.; Mullican, S. E.; Alenghat, T.; Liu, X. S.; Lazar, M. a. A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism. *Science* **2011**, *331*, 1315–9.

[111] Dhalluin, C.; Carlson, J. E.; Zeng, L.; He, C.; Aggarwal, A. K.; Zhou, M. M. Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **1999**, *399*, 491–6.

[112] Lin, H.-Y.; Chen, C.-S.; Lin, S.-P.; Weng, J.-R.; Chen, C.-S. Targeting histone deacetylase in cancer therapy. *Medicinal Research Reviews* **2006**, *26*, 397–413.

[113] Garber, K. HDAC inhibitors overcome first hurdle. *Nature Biotechnology* **2007**, *25*, 17–9.

[114] Glaser, K. B. HDAC inhibitors: clinical update and mechanism-based potential. *Biochemical Pharmacology* **2007**, *74*, 659–71.

[115] Witt, O.; Deubzer, H. E.; Milde, T.; Oehme, I. HDAC family: What are the cancer relevant targets? *Cancer Letters* **2009**, *277*, 8–21.

[116] Marson, C. M. Histone deacetylase inhibitors: design, structure-activity relationships and therapeutic implications for cancer. *Anti-Cancer Agents in Medicinal Chemistry* **2009**, *9*, 661–92.

[117] Minucci, S.; Pelicci, P. G. Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. *Nature Reviews. Cancer* **2006**, *6*, 38–51.

[118] Gurard-Levin, Z. A.; Kilian, K. A.; Kim, J.; Bähr, K.; Mrksich, M. Pep-

tide arrays identify isoform-selective substrates for profiling endogenous lysine deacetylase activity. *ACS Chemical Biology* **2010**, *5*, 863–73.

[119] Wolfson, N. A.; Ann Pitcairn, C.; Fierke, C. A. HDAC8 substrates: Histones and beyond. *Biopolymers* **2013**, *99*, 112–26.

[120] Wilson, B. J.; Tremblay, A. M.; Deblois, G.; Sylvain-Drolet, G.; Giguère, V. An acetylation switch modulates the transcriptional activity of estrogen-related receptor alpha. *Molecular Endocrinology* **2010**, *24*, 1349–58.

[121] Deardorff, M. A. et al. HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. *Nature* **2012**, *489*, 313–7.

[122] Olson, D. E. et al. An unbiased approach to identify endogenous substrates of"histon" deacetylase 8. *ACS Chemical Biology* **2014**, in press.

[123] Kaiser, F. J. et al. Loss-of-function HDAC8 mutations cause a phenotypic spectrum of Cornelia de Lange syndrome-like features, ocular hypertelorism, large fontanelle and X-linked inheritance. *Human Molecular Genetics* **2014**, *23*, 2888–900.

[124] Decroos, C.; Bowman, C. M.; Moser, J.-A. S.; Christianson, K. E.; Deardorff, M. A.; Christianson, D. W. Compromised structure and function of HDAC8 mutants identified in Cornelia de Lange Syndrome spectrum disorders. *ACS Chemical Biology* **2014**, in press.

[125] Somoza, J. R. et al. Structural snapshots of human HDAC8 provide insights into the class I histone deacetylases. *Structure* **2004**, *12*, 1325–34.

[126] Vannini, A.; Volpari, C.; Gallinari, P.; Jones, P.; Mattu, M.; Carfí, A.; De Francesco, R.; Steinkühler, C.; Di Marco, S. Substrate binding to histone deacetylases as shown by the crystal structure of the HDAC8-substrate complex. *EMBO Reports* **2007**, *8*, 879–84.

[127] Dowling, D. P.; Gantt, S. L.; Gattis, S. G.; Fierke, C. A.; Christianson, D. W. Structural studies of human histone deacetylase 8 and its site-specific variants complexed with substrate and inhibitors. *Biochemistry* **2008**, *47*, 13554–63.

[128] Gantt, S. L.; Joseph, C. G.; Fierke, C. A. Activation and inhibition of histone deacetylase 8 by monovalent cations. *The Journal of Biological Chemistry* **2010**, *285*, 6036–43.

[129] Gantt, S.; Gattis, S. G.; Fierke, C. A. Catalytic activity and inhibition of human histone deacetylase 8 is dependent on the identity of the active site metal ion. *Biochemistry* **2006**, *45*, 6170–8.

[130] Watson, P. J.; Fairall, L.; Santos, G. M.; Schwabe, J. W. R. Structure of HDAC3 bound to co-repressor and inositol tetraphosphate. *Nature* **2012**, *481*, 335–40.

[131] Lee, H.; Rezai-Zadeh, N.; Seto, E. Negative regulation of histone deacetylase 8 activity by cyclic AMP-dependent protein kinase A. *Molecular and Cellular Biology* **2004**, *24*, 765–73.

[132] Werbeck, N. D.; Kirkpatrick, J.; Reinstein, J.; Hansen, D. F. Using 15N-ammonium to characterise and map potassium binding sites in proteins by NMR spectroscopy. *Chembiochem : a European journal of chemical biology* **2014**, *15*, 543–8.

[133] Wu, R.; Wang, S.; Zhou, N.; Cao, Z.; Zhang, Y. A proton-shuttle reaction mechanism for histone deacetylase 8 and the catalytic role of metal ions. *Journal of the American Chemical Society* **2010**, *132*, 9471–9.

[134] Chen, K.; Zhang, X.; Wu, Y.-D.; Wiest, O. Inhibition and mechanism of HDAC8 revisited. *Journal of the American Chemical Society* **2014**, *136*, 11636–43.

[135] Villali, J.; Kern, D. Choreographing an enzyme's dance. *Current opinion in chemical biology* **2010**, *14*, 636–43.

[136] Fraser, J. S.; Clarkson, M. W.; Degnan, S. C.; Erion, R.; Kern, D.; Alber, T. Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **2009**, *462*, 669–73.

[137] Bhabha, G.; Lee, J.; Ekiert, D. C.; Gam, J.; Wilson, I. a.; Dyson, H. J.; Benkovic, S. J.; Wright, P. E. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* **2011**, *332*, 234–8.

[138] Rao, J. N.; Jao, C. C.; Hegde, B. G.; Langen, R.; Ulmer, T. S. A combinatorial NMR and EPR approach for evaluating the structural ensemble of partially folded proteins. *Journal of the American Chemical Society* **2010**, *132*, 8657–68.

[139] Uversky, V. N. Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. *Chemical Society Reviews* **2011**, *40*, 1623–34.

[140] Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D. E. Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *Journal of the American Chemical Society* **2012**, *134*, 3787–91.

[141] Xue, Y.; Skrynnikov, N. R. Motion of a disordered polypeptide chain as studied by paramagnetic relaxation enhancements, 15N relaxation, and molecular dynamics simulations: how fast is segmental diffusion in denatured ubiquitin? *Journal of the American Chemical Society* **2011**, *133*, 14614–28.

[142] Vallurupalli, P.; Hansen, D. F.; Stollar, E.; Meirovitch, E.; Kay, L. E. Measurement of bond vector orientations in invisible excited states of

proteins. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 18473–7.

[143] Werbeck, N. D.; Itzhaki, L. S. Probing a moving target with a plastic unfolding intermediate of an ankyrin-repeat protein. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 7863–8.

[144] Papaleo, E.; Sutto, L.; Gervasio, F. L.; Lindorff-Larsen, K. Conformational changes and free energies in a Proline Isomerase. *Journal of Chemical Theory and Computation* **2014**, *10*, 4169–74.

[145] Fersht, A. R.; Sato, S. Phi-value analysis and the nature of protein-folding transition states. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 7976–81.

[146] Fersht, A. R.; Leatherbarrow, R. J.; Wells, T. N. C. Quantitative analysis of structure-activity relationships in engineered proteins by linear free-energy relationships. *Nature* **1986**, *322*, 284–6.

[147] Carter, P. J.; Winter, G.; Wilkinson, A. J.; Fersht, A. R. The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (Bacillus stearothermophilus). *Cell* **1984**, *38*, 835–40.

[148] Kiel, C.; Serrano, L.; Herrmann, C. A detailed thermodynamic analysis of ras/effector complex interfaces. *Journal of Molecular Biology* **2004**, *340*, 1039–58.

[149] Brockwell, D. J.; Radford, S. E. Intermediates: ubiquitous species on folding energy landscapes? *Current Opinion in Structural Biology* **2007**, *17*, 30–7.

[150] Marson, C. M.; Matthews, C. J.; Yiannaki, E.; Atkinson, S. J.; Soden, P. E.; Shukla, L.; Lamadema, N.; Thomas, N. S. B. Discovery

of potent, isoform-selective inhibitors of histone deacetylase containing chiral heterocyclic capping groups and a N-(2-Aminophenyl)benzamide binding unit. *Journal of Medicinal Chemistry* **2013**, *56*, 6156–74.

[151] Wu, R.; Lu, Z.; Cao, Z.; Zhang, Y. Zinc chelation with hydroxamate in histone deacetylases modulated by water access to the linker binding channel. *Journal of the American Chemical Society* **2011**, *133*, 6110–3.

[152] Brunsteiner, M.; Petukhov, P. A. Insights from comprehensive multiple receptor docking to HDAC8. *Journal of Molecular Modeling* **2012**, *18*, 3927–39.

[153] Thangapandian, S.; Sundarapandian, T.; John, S.; Shalini, J.; Sakkiah, S.; Sugunadevi, S.; Lee, K. W.; Woo, L. K. Docking-enabled pharmacophore model for histone deacetylase 8 inhibitors and its application in anti-cancer drug discovery. *Journal of Molecular Graphics & Modelling* **2010**, *29*, 382–95.

[154] Wang, D.-F.; Helquist, P.; Wiech, N. L.; Wiest, O. Toward selective histone deacetylase inhibitor design: homology modeling, docking studies, and molecular dynamics simulations of human class I histone deacetylases. *Journal of Medicinal Chemistry* **2005**, *48*, 6936–47.

[155] Wolfson, N. A.; Pitcairn, C. A.; Sullivan, E. D.; Joseph, C. G.; Fierke, C. A. An enzyme-coupled assay measuring acetate production for profiling histone deacetylase specificity. *Analytical biochemistry* **2014**, *456*, 61–9.

[156] Gurard-Levin, Z. a.; Mrksich, M. The activity of HDAC8 depends on local and distal sequences of its peptide substrates. *Biochemistry* **2008**, *47*, 6242–50.

[157] Fatkins, D. G.; Zheng, W. A spectrophotometric assay for histone deacetylase 8. *Analytical Biochemistry* **2008**, *372*, 82–8.

[158] Joerger, A. C.; Fersht, A. R. Structure-function-rescue: the diverse nature of common p53 cancer mutants. *Oncogene* **2007**, *26*, 2226–42.

[159] Vuzman, D.; Levy, Y. Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Molecular bioSystems* **2012**, *8*, 47–57.

[160] Hansen, D. F.; Yang, D.; Feng, H.; Zhou, Z.; Wiesner, S.; Bai, Y.; Kay, L. E. An exchange-free measure of 15N transverse relaxation: an NMR spectroscopy application to the study of a folding intermediate with pervasive chemical exchange. *Journal of the American Chemical Society* **2007**, *129*, 11468–79.

[161] Baldwin, A. J.; Hansen, D. F.; Vallurupalli, P.; Kay, L. E. Measurement of methyl axis orientations in invisible, excited states of proteins by relaxation dispersion NMR spectroscopy. *Journal of the American Chemical Society* **2009**, *131*, 11939–48.

[162] Hansen, A. L.; Lundström, P.; Velyvis, A.; Kay, L. E. Quantifying millisecond exchange dynamics in proteins by CPMG relaxation dispersion NMR using side-chain 1H probes. *Journal of the American Chemical Society* **2012**, *134*, 3178–89.

[163] Skrynnikov, N. R.; Millet, O.; Kay, L. E. Deuterium spin probes of side-chain dynamics in proteins. 2. Spectral density mapping and identification of nanosecond time-scale side-chain motions. *Journal of the American Chemical Society* **2002**, *124*, 6449–60.

[164] Muhandiram, D. R.; Yamazaki, T.; Sykes, B. D.; Kay, L. E. Measurement of 2H T1 and T1rho relaxation times in uniformly 13C-labeled and fractionally 2H-labeled proteins in solution. *Journal of the American Chemical Society* **1995**, *117*, 11536–44.

[165] Horst, R.; Bertelsen, E. B.; Fiaux, J.; Wider, G.; Horwich, A. L.; Wüthrich, K. Direct NMR observation of a substrate protein bound to

the chaperonin GroEL. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 12748–53.

[166] Kay, L. E.; Bull, T. E. Heteronuclear Transverse Relaxation in AMX, AX2, and AX3 Spin Systems. *Journal of Magnetic Resonance* **1992**, *99*, 615–22.

[167] Kay, L. E.; Keifer, P.; Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *Journal of the American Chemical Society* **1992**, *114*, 10663–5.

[168] Palmer, A. G.; Cavanagh, J.; Wright, P. E.; Rance, M. Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *Journal of Magnetic Resonance* **1991**, *93*, 151–170.

[169] Baldwin, A. J.; Religa, T. L.; Hansen, D. F.; Bouvignies, G.; Kay, L. E. 13CHD2 methyl group probes of millisecond time scale exchange in proteins by 1H relaxation dispersion: an application to proteasome gating residue dynamics. *Journal of the American Chemical Society* **2010**, *132*, 10992–5.

[170] Leigh Jr, J. S. Relaxation times in systems with chemical exchange: some exact solutions. *Journal of Magnetic Resonance* **1971**, *311*, 308–11.

[171] Hansen, D. F.; Led, J. J. Implications of using approximate Bloch-McConnell equations in NMR analyses of chemically exchanging systems: application to the electron self-exchange of plastocyanin. *Journal of Magnetic Resonance* **2003**, *163*, 215–27.

[172] Jacobsen, J.; Bilds{\o}e, H.; Schaumburg, K. Application of density matrix formalism in NMR spectroscopy. II. The one-spin-1 case in anisotropic phase. *Journal of Magnetic Resonance* **1976**, *23*, 153–64.

[173] Mittermaier, A.; Kay, L. E. Measurement of methyl 2H quadrupolar couplings in oriented proteins. How uniform is the quadrupolar coupling constant? *Journal of the American Chemical Society* **1999**, *121*, 10608–13.

[174] Wang, D.-F.; Wiest, O.; Helquist, P.; Lan-Hargest, H.-Y.; Wiech, N. L. On the function of the 14 \AA long internal cavity of histone deacetylase-like protein: implications for the design of histone deacetylase inhibitors. *Journal of Medicinal Chemistry* **2004**, *47*, 3409–17.

[175] Wambua, M. K.; Nalawansha, D. A.; Negmeldin, A. T.; Pflum, M. K. H. Mutagenesis studies of the 14 \AA internal cavity of histone deacetylase 1: insights toward the acetate-escape hypothesis and selective inhibitor design. *Journal of Medicinal Chemistry* **2014**, *57*, 642–50.

[176] Whitehead, L. et al. Human HDAC isoform selectivity achieved via exploitation of the acetate release channel with structurally unique small molecule inhibitors. *Bioorganic & Medicinal Chemistry* **2011**, *19*, 4626–34.

[177] Smart, O. S.; Goodfellow, J. M.; Wallace, B. A. The pore dimensions of gramicidin A. *Biophysical Journal* **1993**, *65*, 2455–60.

[178] George, S. J.; Lowery, M. D.; Solomon, E. I.; Cramer, S. P. Copper L-edge spectral studies: a direct experimental probe of the ground-state covalency in the blue copper site in plastocyanin. *Journal of the American Chemical Society* **1993**, *115*, 2968–2969.

[179] Strange, R. W.; Ellis, M.; Hasnain, S. Atomic resolution crystallography and XAFS. *Coordination Chemistry Reviews* **2005**, *249*, 197–208.

[180] Penfield, K. W.; Gewirth, A. A.; Solomon, E. I. Electronic structure and bonding of the blue copper site in plastocyanin. *Journal of the American Chemical Society* **1985**, *107*, 4519–4529.

[181] Wilkens, S. J.; Xia, B.; Volkman, B. F.; Weinhold, F.; Markley, J. L.; Westler, W. M. Inadequacies of the point-dipole approximation for describing electron-nuclear interactions in paramagnetic proteins: hybrid density functional calculations and the analysis of NMR relaxation of high-spin iron(III) rubredoxin. *The Journal of Physical Chemistry B* **1998**, *102*, 8300–5.

[182] Hansen, D. F.; Led, J. J. Mapping the electronic structure of the blue copper site in plastocyanin by NMR relaxation. *Journal of the American Chemical Society* **2004**, *126*, 1247–52.

[183] Led, J. J.; Neesgaard, E.; Johansen, J. T. Carbon dioxide hydration activity and metal-substrate distances of manganese (II) human carbonic anhydrase B determined by 13C magnetization-transfer NMR. *FEBS Letters* **1982**, *147*, 74–80.

[184] Gueron, M. Nuclear relaxation in macromolecules by paramagnetic ions: a novel mechanism. *Journal of Magnetic Resonance* **1975**, *19*, 58–66.

[185] Volkman, B. F.; Wilkens, S. J.; Lee, A. L.; Xia, B.; Westler, W. M.; Beger, R.; Markley, J. L. Redox-Dependent Magnetic Alignment of Clostridium pasteurianum Rubredoxin: Measurement of Magnetic Susceptibility Anisotropy and Prediction of Pseudocontact Shift Contributions. *Journal of the American Chemical Society* **1999**, *121*, 4677–4683.

[186] Lin, I.-J.; Xia, B.; King, D. S.; Machonkin, T. E.; Westler, W. M.; Markley, J. L. Hyperfine-shifted (13)C and (15)N NMR signals from Clostridium pasteurianum rubredoxin: extensive assignments and quantum chemical verification. *Journal of the American Chemical Society* **2009**, *131*, 15555–63.

[187] Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Al-

gorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* **2008**, *4*, 435–47.

[188] Feynman, R. Simulating physics with computers. *International Journal of Theoretical Physics* **1982**, *21*, 67–488.

[189] Kane, B. E. A silicon-based nuclear spin quantum computer. *Nature* **1998**, *393*, 133–137.

[190] Simon, D. R. On the power of quantum computation. 1994.

[191] Grinolds, M. S.; Warner, M.; De Greve, K.; Dovzhenko, Y.; Thiel, L.; Walsworth, R. L.; Hong, S.; Maletinsky, P.; Yacoby, A. Subnanometre resolution in three-dimensional magnetic resonance imaging of individual dark spins. *Nature Nanotechnology* **2014**, *9*, 279–84.

[192] Yang, W.; Liu, R.-B. Quantum many-body theory of qubit decoherence in a finite-size spin bath. *Physical Review B* **2008**, *78*, 85315.

[193] Mohammady, M. H.; Morley, G. W.; Monteiro, T. S. Bismuth Qubits in Silicon: The Role of EPR Cancellation Resonances. *Physical Review Letters* **2010**, *105*, 067602.

[194] Belli, M.; Fanciulli, M.; Abrosimov, N. V. Pulse electron spin resonance investigation of bismuth-doped silicon: Relaxation and electron spin echo envelope modulation. *Physical Review B* **2011**, *83*, 235204.

[195] Medvedev, Z. A. An attempt at a rational classification of theories of ageing. *Biological Reviews* **1990**, *65*, 375–398.

[196] Sohal, R. S.; Weindruch, R. Oxidative Stress, Caloric Restriction, and Aging. *Science* **1996**, *273*, 59–63.

[197] Pérez, V. I.; Bokov, A.; Van Remmen, H.; Mele, J.; Ran, Q.; Ikeno, Y.; Richardson, A. Is the oxidative stress theory of aging dead? *Biochimica et biophysica acta* **2009**, *1790*, 1005–14.

[198] Van Raamsdonk, J. M.; Hekimi, S. Deletion of the mitochondrial superoxide dismutase sod-2 extends lifespan in Caenorhabditis elegans. *PLoS Genetics* **2009**, *5*, e1000361.

[199] Gems, D.; Doonan, R. Antioxidant defense and aging in C. elegans : Is the oxidative damage theory of aging wrong? *Cell Cycle* **2014**, *8*, 1681–1687.

[200] Meneghini, R. Iron homeostasis, oxidative stress, and DNA damage. *Free Radical Biology and Medicine* **1997**, *23*, 783–792.

[201] Polla, A. S.; Polla, L. L.; Polla, B. S. Iron as the malignant spirit in successful ageing. *Ageing Research Reviews* **2003**, *2*, 25–37.

[202] Kim, Y.-I.; Cho, J. H.; Yoo, O. J.; Ahnn, J. Transcriptional regulation and life-span modulation of cytosolic aconitase and ferritin genes in C.elegans. *Journal of molecular biology* **2004**, *342*, 421–33.

[203] Pate, K. T.; Rangel, N. A.; Fraser, B.; Clement, M. H. S.; Srinivasan, C. Measuring "free" iron levels in Caenorhabditis elegans using low-temperature Fe(III) electron paramagnetic resonance spectroscopy. *Analytical Viochemistry* **2006**, *358*, 199–207.

[204] Bressi, J. C.; Jennings, A. J.; Skene, R.; Wu, Y.; Melkus, R.; De Jong, R.; O'Connell, S.; Grimshaw, C. E.; Navre, M.; Gangloff, A. R. Exploration of the HDAC2 foot pocket: Synthesis and SAR of substituted N-(2-aminophenyl) benzamides. *Bioorganic & Medicinal Chemistry Letters* **2010**, *20*, 3142–5.

[205] Jarzynski, C. Nonequilibrium equality for free energy differences. *Physical Review Letters* **1997**, *78*, 2690–3.

[206] Kannan, A.; Camilloni, C.; Sahakyan, A. B.; Cavalli, A.; Vendruscolo, M. A conformational ensemble derived using NMR methyl chemical shifts

reveals a mechanical clamping transition that gates the binding of the HU protein to DNA. *Journal of the American Chemical Society* **2014**, *136*, 2204–7.

[207] Boomsma, W.; Tian, P.; Frellsen, J.; Ferkinghoff-Borg, J.; Hamelryck, T.; Lindorff-Larsen, K.; Vendruscolo, M. Equilibrium simulations of proteins using molecular fragment replacement and NMR chemical shifts. *Proceedings of the National Academy of Sciences* **2014**, *111*, 13852–7.

[208] Duarte, F.; Bauer, P.; Barrozo, A.; Amrein, B. A.; Purg, M.; Aqvist, J.; Kamerlin, S. C. L. Force field independent metal parameters using a nonbonded dummy model. *The journal of physical chemistry. B* **2014**, *118*, 4351–62.

[209] Michaelis, L.; Menten, M. L. Die kinetik der invertinwirkung. *Biochemie Z* **1913**, *49*, 352.

[210] Michaelis, L.; Menten, M. L.; Johnson, K. a.; Goody, R. S. The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. *Biochemistry* **2011**, *50*, 8264–9.

[211] Hwang, T.; Shaka, A. Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *Journal of Magnetic Resonance, Series A* **1995**, *112*, 275–279.

[212] Guerois, R.; Nielsen, J. E.; Serrano, L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *Journal of Molecular Biology* **2002**, *320*, 369–87.

[213] Lindorff-Larsen, K.; Paci, E.; Serrano, L.; Dobson, C. M.; Vendruscolo, M. Calculation of mutational free energy changes in transition states for protein folding. *Biophysical Journal* **2003**, *85*, 1207–14.

[214] Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: a

message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **1995**, *91*, 43–56.

[215] van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: fast, flexible, and free. *Journal of Computational Chemistry* **2005**, *26*, 1701–18.

[216] Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **2001**, *7*, 306–17.

[217] Feig, M.; Karanicolas, J.; Brooks, C. L. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *Journal of Molecular Graphics & Modelling* **2004**, *22*, 377–95.

[218] Otting, G. Protein NMR using paramagnetic ions. *Annual Review of Biophysics* **2010**, *39*, 387–405.

[219] Kristjansdottir, S.; Lindorff-Larsen, K.; Fieber, W.; Dobson, C. M.; Vendruscolo, M.; Poulsen, F. M. Formation of native and non-native interactions in ensembles of denatured ACBP molecules from paramagnetic relaxation enhancement studies. *Journal of Molecular Biology* **2005**, *347*, 1053–62.

[220] Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–8.

[221] Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **2011**, *32*, 2319–27.

[222] Sali, A.; Blundell, T. Comparative protein modelling by satisfaction of spatial restraints. *Protein Structure by Distance Analysis* **1994**, 64–86.

[223] Neese, F. The ORCA program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *2*, 73–78.

[224] Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Research* **2005**, *33*, W368–71.

[225] Hasegawa, H.; Holm, L. Advances and pitfalls of protein structural alignment. *Current Opinion in Structural Biology* **2009**, *19*, 341–8.

[226] Frishman, D.; Argos, P. Knowledge-based protein secondary structure assignment. *Proteins* **1995**, *23*, 566–79.

A. Appendix

A.1. Sequence Alignment of Class I HDACs

Figure A.1.: Sequence alignment of human class I HDACs (1,2,3 and 8) as done by Clustal Omega (EMBL-EBI). Colour indicate type of residue and sequence conservations is indicated by signs under the alignment (* : . - descending in this order).

A.2. ^{13}CD Scalar Coupling

For simplicity the normalisation of the operators is neglected as here only the qualitative result is important. In the following I calculate the commutator of the centre line (Formula 2.1) and inner line (Formula 2.3) of the ^{13}CD spin system with the scalar coupling Hamiltonian using the lowering and raising operator representation.

Central line:

$$
\begin{aligned}
&[C_+(1-D_z^2), C_zD_z] \\
&= \underbrace{[C_+, C_z]D_z}_{=-C_+D_z} + \underbrace{C_z[C_+, D_z]}_{=0} - [C_+D_z^2, C_zD_z] \\
&= -C_+D_z - C_+D_z \underbrace{[D_z, C_zD_z]}_{=0} - C_+ \underbrace{[D_z, C_zD_z]}_{=0} D_z - \underbrace{[C_+, C_zD_z]}_{=-C_+D_z} D_z^2 \\
&= -C_+D_z + C_+D_z^3 \\
&= 0\,,
\end{aligned}
$$

since $D_z^3 = D_z$, confirming that the central line does not undergo scalar coupling.

Inner line:

$$
\begin{aligned}
&[C_+(D_z^2 - D_z), C_z D_z] \\
&= [C_+ D_z^2, C_z D_z] - [C_+ D_z, C_z D_z] \\
&= C_+ D_z \underbrace{[D_z, D_z C_z]}_{=0} + C_+ \underbrace{[D_z, C_z D_z]}_{=0} D_z + \underbrace{[C_x, C_z D_z] D_z^2}_{=-C_+ D_z^3} - [C_x D_z, C_z D_z] \\
&= -C_+ D_z^3 - C_+ \underbrace{[D_z, C_z]}_{=0} D_z - C_+ C_z \underbrace{[D_z, D_z]}_{=0} - \underbrace{[C_+, C_z]}_{=-C_+ D_z^2} - C_z \underbrace{[C_+, D_z]}_{=0} D_z \\
&= -C_+ D_z^3 + C_+ D_z^2 \\
&= C_+(D_z^2 - D_z)
\end{aligned}
$$

Hence $C_+(D_z^2 - D_z)$ is an eigenfunction and evolves under scalar coupling. By analogy the same behaviour can be inferred for the outer line.

A.3. $^{13}CH_2D$ TROSY Post-Processing

The following presents a script that automates the post-processing of a $^{13}CH_2D$ TROSY. This script calls other scripts that use NMRpipe to shift spectra etc. these scripts are shown later. It also relies on the data being 'de-scrambled', i.e. the T_1 increments being in the correct order as they are usually 'scrambled' to remove any trends from the relaxation measurements (Section 5.3.1).

Process wrapper script process.csh:

```
1  #!/bin/tcsh
2
3  set counter=0
4  set n_t1=8
5
6  foreach plane ( 1 3 5 7 9 11 13 15 )
7      @ counter++
8      #echo $counter
9      echo $plane
10     set data_base='data/test';
11     set next_plane=`expr $plane + 1`;
12     set ipapa=`printf "%s%03d%s" $data_base $plane ".ft2"`;
13     set ipapb=`printf "%s%03d%s" $data_base $next_plane ".ft2"`;
14     set final_plane=`printf "%s%03d%s" "plane" $counter ".ft2"`;
15     set final_plane_neg=`printf "%s%03d%s" "plane" $counter "_neg.
           ft2"`;
16     #echo $plane
17     #echo $next_plane
18     echo $counter
19
20     #SUM
21     addNMR -in1 $ipapa -in2 $ipapb -add -out sum.ft2
22
23     #DIFF
24     addNMR -in1 $ipapa -in2 $ipapb -sub -out diff.ft2
25
```

```
26      #90 phase
27      ./sum2ps90.com
28      # SUM and DIFF
29      addNMR -in1 sum90.ft2 -in2 diff.ft2 -add -out seven.ft2
30      addNMR -in1 diff.ft2 -in2 sum90.ft2 -sub -out eight.ft2
31
32      #shift
33      ./eight_leftshift.com
34
35      #SUM
36      addNMR -in1 eight_left.ft2 -in2 seven.ft2 -add -out
            $final_plane
37      addNMR -in1 eight_left.ft2 -in2 seven.ft2 -sub -out
            $final_plane_neg
38
39      #change headers to make cube of nT/2
40      sethdr  $final_plane -zN $n_t1 -zT $n_t1
41      sethdr  $final_plane_neg -zN $n_t1 -zT $n_t1
42
43  end
44  #create the cube
45  xyz2pipe -in plane%03d.ft2 | nmrPipe -out final_cube.ft2 -ov -verb
46  xyz2pipe -in plane%03d_neg.ft2 | nmrPipe -out final_cube_neg.ft2 -
        ov -verb
47  #make first plane 2D for CCPN read-in
48  #make sure its the plane with best S/N for peak picking
49  cp plane001.ft2 plane001_ccpn.ft2
50  sethdr plane001_ccpn.ft2 -ndim 2 -nT 1 -nZ 1
51  #EOF
```

Phase shift the summed up spectrum with sum2ps90.com:

```
1  #!/bin/csh
2
3  nmrPipe -verb -in sum.ft2                                    \
4      |    nmrPipe -fn HT  \
```

```
5     |     nmrPipe -fn PS -p0 90. -di                                \
6     |     nmrPipe -ov -verb -out sum90.ft2
```

Circular shift script eight_leftshift.com:

```
1 #!/bin/csh
2
3 nmrPipe -verb -in eight.ft2                                         \
4     | nmrPipe -fn CS -ls 124Hz \
5     |     nmrPipe -ov -verb -out eight_left.ft2
```

A.4. Water Tunnel Analysis Code

After defining the three boxes to define the transient tunnel the water molecules are assigned as being outside of all boxes or in one of the three and labelled accordingly, i.e. 0, 1, 2 or 3. This is achieved by running a set of loops that check the distance of each oxygen atom to each side (plane) of the boxes. If the absolute distance is smaller than the distance to the parallel side for each box the molecule is labeled accordingly. It should be noted that the presented code only works for 6-sided boxes with at least triclinic symmetry, i.e. 6-sided box defined by 3 basis vectors.

```
%% box sort
% caclulate which water is in which or no box
h = waitbar(0,'Time for a coffee? ETA unknown');
clock=tic;
oxy_state=zeros(length(time_vec),total_water); %delete states

time_vec=trj(:,1);
for frame=1:length(time_vec)
    % run through oxygens
    for oxy=1:total_water
        posi=trj(frame,oxy*3-1:oxy*3-1+2);
        % run through boxes
        for box_id=1:nr_of_boxes
            checker=0;
            for i=1:6
                distance=abs(dot(box{box_id}.normal{i},(posi-box{
                    box_id}.plane_point{i})));
                if distance > box{box_id}.dist{i}
                    break
                elseif distance <= box{box_id}.dist{i}
                    checker=checker+1;
                end
                if checker==6
```

```
23                          oxy_state(frame,oxy)=box_id;
24                      end
25
26                  end
27              end
28
29          end
30          elapsed=toc(clock); %per frame time
31          waitbar((frame-1)/length(time_vec),h,sprintf('Time for a
                coffee? ETA %2.2f min',(elapsed/60)/frame*(length(time_vec)
                -frame)));
32  end %sort waters into boxes (or no boxes)
33  close(h)
```

The resulting box state trajectories are subsequently analyed for 1-2-3 or 3-2-1-type transitions. The following is an example for the 1-2-3 (out) transition:

```
1  %%
2  % look for transitions 1-2-3 OR 3-2-1
3  in_counter=0;
4  out_counter=0;
5
6  h = waitbar(0,'Time for a coffee? ETA unknown');
7  clock=tic;
8  tmp_trans_vec=[];
9
10 for id=1:total_water %tunnel count start
11     test_time=zeros(length(time_vec),1);
12
13     for time=1:length(time_vec)
14
15         if test_time(time)==0
16             time_tr=time;
17             while oxy_state(time_tr,id)==1 %out count 1-2-3
18                 tmp_trans_vec=[tmp_trans_vec; time_tr oxy_state(
                       time_tr,id) id];
```

```
19                  time_tr=time_tr+1;
20                  if time_tr>length(time_vec)
21                      break
22                  end
23
24                  while oxy_state(time_tr,id)==1 % do as long as in
                        box 1
25                      tmp_trans_vec=[tmp_trans_vec; time_tr
                            oxy_state(time_tr,id) id];
26                      time_tr=time_tr+1;
27                      if time_tr>length(time_vec)
28                          break
29                      end
30                  end
31
32                  if time_tr>length(time_vec)
33                      break
34                  end
35
36                  if oxy_state(time_tr,id)==0
37                      tmp_trans_vec=[];
38                      break
39                  end
40
41                  if oxy_state(time_tr,id)==2
42
43                      while oxy_state(time_tr,id)==2 % do as long as
                            in box 2
44                          tmp_trans_vec=[tmp_trans_vec; time_tr
                                oxy_state(time_tr,id) id];
45                          time_tr=time_tr+1;
46                          if time_tr>length(time_vec)
47                              break
48                          end
49                      end
```

```
50
51                      if time_tr>length(time_vec)
52                          break
53                      end
54                  end
55
56                  if oxy_state(time_tr,id)==0
57                      tmp_trans_vec=[];
58                      break
59                  end
60
61                  if oxy_state(time_tr,id)==3
62                      out_counter=out_counter+1;
63                      tmp_trans_vec=[tmp_trans_vec; time_tr
                            oxy_state(time_tr,id) id];
64                      out_trans{out_counter}=tmp_trans_vec;
65                      tmp_trans_vec=[];
66                      test_time(1:time_tr)=1;
67                      time_tr=time_tr+1;
68                      break
69                  end
70
71                  if time_tr>length(time_vec)
72                      break
73                  end
74
75
76              end
77          end
78          if time_tr>length(time_vec)
79              break
80          else
81              test_time(1:time_tr)=1;
82          end
83      end
```

Within the outer loop the 3-2-1 type transitions are analysed as well, but cut off here for the sake of brevity.

A.5. Order Parameter Fitting

The following script was written to extract order parameters for an LS2 model from deuterium relaxation data.

```
%% fitting order parameteres (S_f2) for CH2D groups using transverse and longitudinal
% relaxation rates of deuterium assumming an LS2 model - mod for 4 RATES
% literature
% Millet, OD; Muhandiram, DR; Skrynnikov, NR and Kay, LE;
% "Deuterium Spin Probes of Side-Chain Dynamcis in Proteins 1+2"
% 2012 JACS
% start:
clear all;
close all;
sample_name='hdac8_wt_ile_ch2d';
% run settings
run_on_mac=0;          % dropbox flag
fit_type=1;            % 1/0 MC or no MC for error est.
sim_data=1;            % compare with simulated data
exclude_peaks=[45 11];  % which peaks to exclude/dummy peaks for fuda fitting etc.
mc_it=100;              % MC iterations
omega_d=92.1*10^(6)*2*pi; % 600MHz spectrometer
assign_vec=[34 56 45 365 94 -1 135 369 322 127 120 172 291 362 284 331 300 231 243 162 19 269 348] ;

if run_on_mac==1
    sim_data=0
end
% variables and constants
phi=109.5*(pi/180); % angle of C-D bond and methyl symmetry axis
tau_f= 20*10^(-12); % correlation time of the fast internal motions
```

```
26 tau_r= 35*10^(-9); % assumed isotropic correlation tumbling time
       of the protein
27 tau=1/(tau_r^(-1)+tau_f^(-1));
28 alpha=(3*cos(phi)^2-1)^2/4;
29 q_coupl= 167*10^(3)*2*pi; %quadru. coupling constant e^2 q Q^2/h -
       taken from lit. - BEWARE of cgs units in the lit.
30 S_f2=0.8 ; % dummy value to start fit
31 ile_vec=[19 34 45 56 94 108 115 120 127 135 162 172  231 235 243
       269 284 291 300 322 331 348 362 365 369];
32
33 % general functions
34 % spectral density model
35 J_fit=@(S_f2,omega,tau) alpha*S_f2*tau_r./(1+(omega.*tau_r).^2)
       +(1-alpha*S_f2)*tau./(1+(omega.*tau).^2); %LS2 model
36 % quadrup. deuterium relaxation rates - 4 RATES
37 R_Dz_fit=@(S_f2,tau) (3/40)*q_coupl^2*(J_fit(S_f2,omega_d,tau)+4*
       J_fit(S_f2,2*omega_d,tau));
38 R_Dplus_fit=@(S_f2,tau) (1/80)*q_coupl^2*(9*J_fit(S_f2,0,tau) +
       15*J_fit(S_f2,omega_d,tau) + 6*J_fit(S_f2,2*omega_d,tau));
39 R_3Dz_2_fit=@(S_f2,tau) (3/40)*q_coupl^2*3*J_fit(S_f2,2*omega_d,
       tau);
40 R_DplusDz_fit=@(S_f2,tau) (1/80)*q_coupl^2*(9*J_fit(S_f2,0,tau)+3*
       J_fit(S_f2,omega_d,tau)+6*J_fit(S_f2,2*omega_d,tau));
41
42
43 %load data
44 if run_on_mac==1
45
46     % load R1's
47     cd('/Users/micha/Dropbox/NMR_data/600_oxford/107/out/');
48     R_Dz=importdata('singleexp.fit',' ',1);
49     % load R2's
50     cd('/Users/micha/Dropbox/NMR_data/600_oxford/110/out/');
51     R_Dplus=importdata('singleexp.fit',' ',1);
52     % load R3Dz...
```

```
53      cd('/Users/micha/Dropbox/NMR_data/600_oxford/109/out/');
54      R_3Dz_2=importdata('singleexp.fit',' ',1);
55      % load RDplusDz....
56      cd('/Users/micha/Dropbox/NMR_data/600_oxford/108/out/');
57      R_DplusDz=importdata('singleexp.fit',' ',1);
58
59  else
60      % load data
61      % load R1's
62      cd('/home/micha/Dropbox/NMR_data/600_oxford/107/out/');
63      R_Dz=importdata('singleexp.fit',' ',1);
64      % load R2's
65      cd('/home/micha/Dropbox/NMR_data/600_oxford/110/out/');
66      R_Dplus=importdata('singleexp.fit',' ',1);
67      % load R1's
68      cd('/home/micha/Dropbox/NMR_data/600_oxford/109/out/');
69      R_3Dz_2=importdata('singleexp.fit',' ',1);
70      % load R2's
71      cd('/home/micha/Dropbox/NMR_data/600_oxford/108/out/');
72      R_DplusDz=importdata('singleexp.fit',' ',1);
73  end
74
75  % sort data for peak alignment (ccpn/fuda can mess them up)
76  % this might need extra attention when some peaks are not assigned
        in some
77  % spectra etc - easy to solve sith dictionary like aligment vector
78  R_Dz.data=sortrows(R_Dz.data,1);
79  R_DplusDz.data=sortrows(R_DplusDz.data,1);
80  R_Dplus.data=sortrows(R_Dplus.data,1);
81  R_3Dz_2.data=sortrows(R_3Dz_2.data,1);
82
83  %% Get peak list and get selected peaks
84  peak_list=R_Dplus.data(:,1);
85  select=1:length(peak_list); % get index of ALL peaks
86  for exc=1:length(exclude_peaks)
```

```
87      excluder=find(peak_list==exclude_peaks(exc)); % find index of
            peak to exclude
88      select=select(select~=excluder); % remove from index vector
89  end
90  store_array=zeros(length(select),4);
91  store_array_single=zeros(length(select),4);
92  store_array_double=zeros(length(select),4);
93  %%
94  save_counter=0;
95  h = waitbar(0,'Time for a coffee? ETA unknown');
96
97  for peak=select
98      save_counter=save_counter+1; % used for saving AND waitbar
99      % get mu and sigma fr each rate
100     R_Dz_obs=R_Dz.data(peak,4);
101     R_Dz_err=R_Dz.data(peak,5);
102     R_Dplus_obs=R_Dplus.data(peak,4);
103     R_Dplus_err=R_Dplus.data(peak,5);
104     R_DplusDz_obs=R_DplusDz.data(peak,4);
105     R_DplusDz_err=R_DplusDz.data(peak,5);
106     R_3Dz_2_obs=R_3Dz_2.data(peak,4);
107     R_3Dz_2_err=R_3Dz_2.data(peak,5);
108
109     % MC loop
110     mc_tmp=zeros(mc_it,2); %pre-allocate for mc data
111     mc_tmp_single=zeros(mc_it,2); %pre-allocate for mc data
112     mc_tmp_double=zeros(mc_it,2); %pre-allocate for mc data
113     clock=tic;
114     for mc=1:mc_it % MC loop
115
116         waitbar((((save_counter-1)*mc_it)+mc)/(length(select)*
                mc_it))
117         %define NEW rates from nomrnd distr.
118         R_Dz_mc=normrnd(R_Dz_obs,R_Dz_err);
119         R_Dplus_mc=normrnd(R_Dplus_obs,R_Dplus_err);
```

```
120         R_DplusDz_mc=normrnd(R_DplusDz_obs,R_DplusDz_err);
121         R_3Dz_2_mc=normrnd(R_3Dz_2_obs,R_3Dz_2_err);
122
123         if fit_type==0 % fit-type clause to use MC or not 1/0
124             % functions dep on S_f2
125             x2_J_mc=@(S_f2) (R_Dz_fit(S_f2,tau)-R_Dz_mc).^2/
                    R_Dz_err.^2+(R_Dplus_fit(S_f2,tau)-R_Dplus_mc).^2/
                    R_Dplus_err.^2+(R_3Dz_2_fit(S_f2,tau)-R_3Dz_2_mc)
                    .^2/R_3Dz_2_err.^2+(R_DplusDz_fit(S_f2,tau)-
                    R_DplusDz_mc).^2/R_DplusDz_err.^2; % err func for
                    MC sim
126             x2_J_mc_single=@(S_f2) (R_Dz_fit(S_f2,tau)-R_Dz_mc)
                    .^2/R_Dz_err.^2+(R_Dplus_fit(S_f2,tau)-R_Dplus_mc)
                    .^2/R_Dplus_err.^2;
127             x2_J_mc_double=@(S_f2) (R_3Dz_2_fit(S_f2,tau)-
                    R_3Dz_2_mc).^2/R_3Dz_2_err.^2+(R_DplusDz_fit(S_f2,
                    tau)-R_DplusDz_mc).^2/R_DplusDz_err.^2;
128
129             % only S_f2 fit
130             mc_tmp(mc,1)=fminsearch(x2_J_mc,0.1,[0.1,0.99]);
131             mc_rmp(mc,2)=tau;
132             mc_tmp_single(mc,1)=fminsearch(x2_J_mc_single
                    ,0.1,[0.1,0.99]);
133             mc_rmp_single(mc,2)=tau;
134             mc_tmp_double(mc,1)=fminsearch(x2_J_mc_double
                    ,0.1,[0.1,0.99]);
135             mc_rmp_double(mc,2)=tau;
136
137         else
138             %functions dep. on S_f2 and tau
139             x2_J_var_mc=@(var) (R_Dz_fit(var(1),var(2))-R_Dz_mc)
                    .^2/R_Dz_err.^2+(R_Dplus_fit(var(1),var(2))-
                    R_Dplus_mc).^2/R_Dplus_err.^2+(R_3Dz_2_fit(var(1),
                    var(2))-R_3Dz_2_mc).^2/R_3Dz_2_err.^2+(
```

```
                R_DplusDz_fit(var(1),var(2))-R_DplusDz_mc).^2/
                R_DplusDz_err.^2;
140             x2_J_var_mc_single=@(var) (R_Dz_fit(var(1),var(2))-
                R_Dz_mc).^2/R_Dz_err.^2+(R_Dplus_fit(var(1),var(2))
                -R_Dplus_mc).^2/R_Dplus_err.^2;
141             x2_J_var_mc_double=@(var) (R_3Dz_2_fit(var(1),var(2))-
                R_3Dz_2_mc).^2/R_3Dz_2_err.^2+(R_DplusDz_fit(var(1)
                ,var(2))-R_DplusDz_mc).^2/R_DplusDz_err.^2;
142
143             % fit S_f2 AND tau
144             var(1)=S_f2;
145             var(2)=tau;
146             var0=[var(1) var(2)];
147             mc_tmp(mc,:)=fminsearch(x2_J_var_mc,var0);
148             mc_tmp_single(mc,:)=fminsearch(x2_J_var_mc_single,var0
                );
149             mc_tmp_double(mc,:)=fminsearch(x2_J_var_mc_double,var0
                );
150         end % end fit-type clause
151
152     end % MC loop end
153
154     ETA=toc(clock);
155     waitbar((((save_counter-1)*mc_it)+mc)/(length(select)*mc_it),h
        ,sprintf('Time for a coffee? ETA %2.2f min',ETA/60*(length(
        select)-save_counter)))
156     S_f2_mu=mean(mc_tmp(:,1));
157     S_f2_sigma=std(mc_tmp(:,1));
158     tau=mean(mc_tmp(:,2));
159
160     S_f2_mu_single=mean(mc_tmp_single(:,1));
161     S_f2_sigma_single=std(mc_tmp_single(:,1));
162     tau_single=mean(mc_tmp_single(:,2));
163
164     S_f2_mu_double=mean(mc_tmp_double(:,1));
```

```
165     S_f2_sigma_double=std(mc_tmp_double(:,1));
166     tau_double=mean(mc_tmp_double(:,2));
167
168     % store in store_arry
169     % all rates
170     store_array(save_counter,1)=peak_list(peak);
171     store_array(save_counter,2)=S_f2_mu;
172     store_array(save_counter,3)=S_f2_sigma;
173     store_array(save_counter,4)=tau;
174     % single
175     store_array_single(save_counter,1)=peak_list(peak);
176     store_array_single(save_counter,2)=S_f2_mu_single;
177     store_array_single(save_counter,3)=S_f2_sigma_single;
178     store_array_single(save_counter,4)=tau_single;
179     % double
180     store_array_double(save_counter,1)=peak_list(peak);
181     store_array_double(save_counter,2)=S_f2_mu_double;
182     store_array_double(save_counter,3)=S_f2_sigma_double;
183     store_array_double(save_counter,4)=tau_double;
184
185 end
186 close(h)
```

A.6. Supplemental MD Results

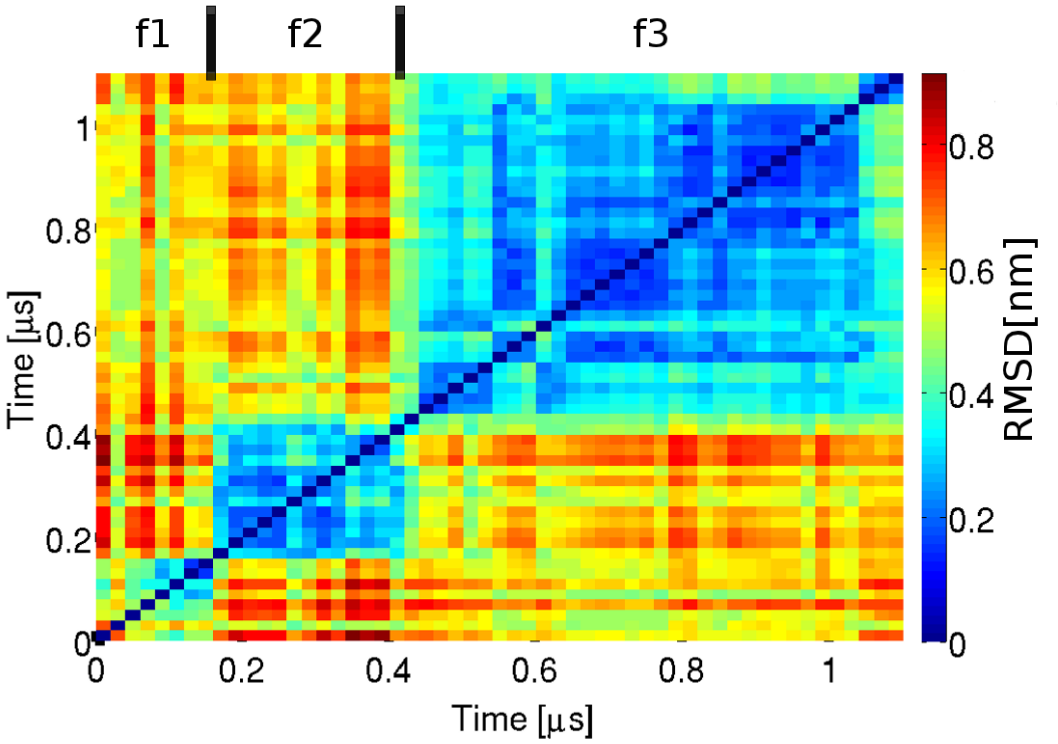


Figure A.2.: Root mean square deviation (RMSD) matrix over the trajectory of the distal L2 loop (residues 83 to 93) and the L1 loop (residues 31 to 35). RMSD calculated for all heavy atoms after removing translational and rotational motion of the protein using a least square fit for the rigid part of the protein (residues with stable secondary structure).

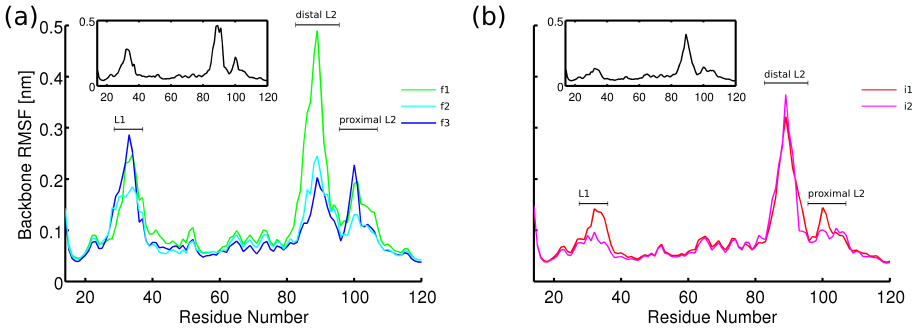


Figure A.3.: Root mean square fluctuation (RMSF) of the L1 and L2 loop region, calculated against the average structure of the protein within the trajectory part of interest. (a): in the three structurally different states f1, f2 and f3 as shown in Figure 2.2(b) in the main text. L1 and L2 loops are highlighted. The inset shows the RMSF over the whole trajectory. From this plot it becomes obvious that the L1 loop does behave similarly in f1 and f3, where it shows a high mobility. Whereas in f2 this loop has decreased mobility due to the salt-bridge formation of Lys33 and the triple Asp repeat (residues 87 to 89) in the L2 loop. The L2 loop is mobile in the f1 state, as expected. In the f2 state it shows decreased mobility due to the aforementioned interactions with the L1 loop via Lys33. In contrast to the L1 loop behaviour this loop retains decreased mobility in the f3 state, which is explained by partial helix formation within the L2 loop (residues 93 to 97). Interestingly, the RMSF of the binding rail residues is similar for the f1 and f3 state, where it shows decreased mobility in the f2, which reflects the trapping of these residues due to the L1:L2 interactions. (b): RMSF plot for the HDAC8:SAHA simulation for the i1 and i2 states. Inset shows the RMSF for the whole trajectory. The L1 loop fluctuations are decreased overall compared to the free form. Moreover, the proximal part of L2 loop exhibits less fluctuation than in the free form, especially the binding rail residues.

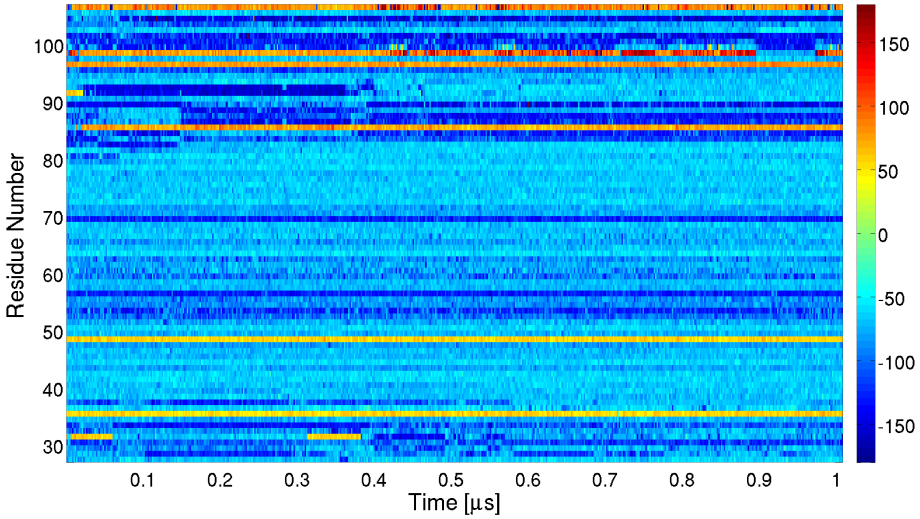


Figure A.4.: Backbone angle ϕ of the L2 loop region shown for the free form simulations. The binding rail flips can be seen at the hinge (Tyr100), as well as the cooperativity with Asp101.

A.7. Supplemental HDAC8 Data

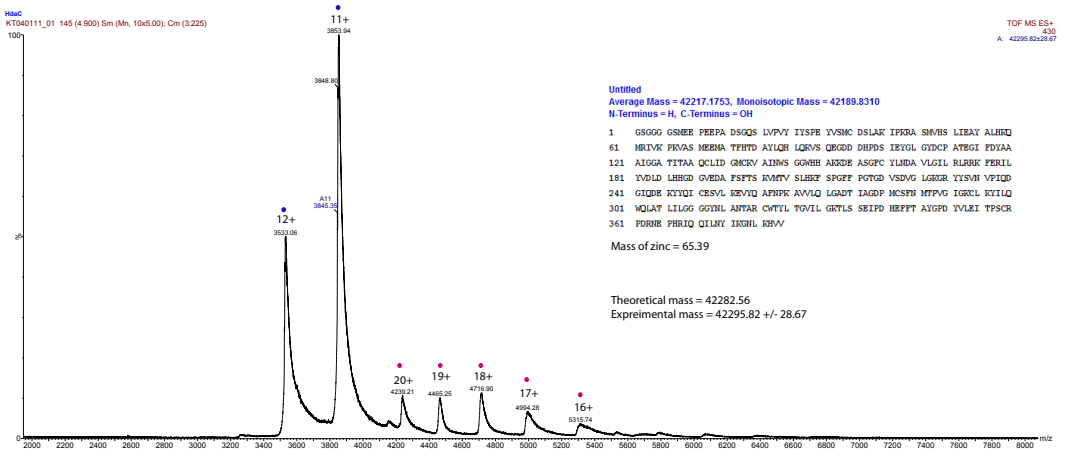


Figure A.5.: MS spectrum of HDAC8 after purification on a S75 gel-filtration column as described in 5.1.1

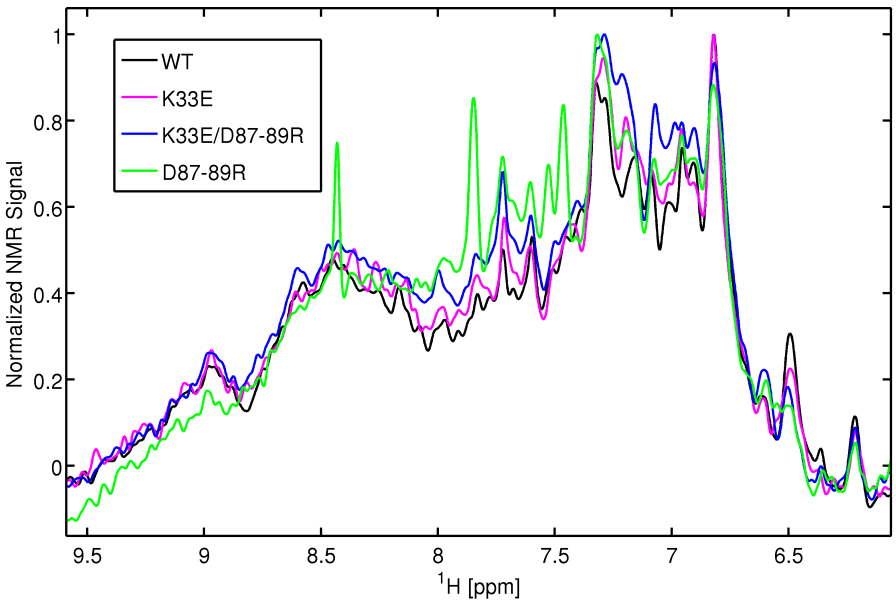


Figure A.6.: Proton NMR spectra overlay of the amide region of a ^{1}H NRM spectra of HDAC8 mutants. HDAC8 WT is shown in black, Lys33Glu mutant is shown in magenta, Asp87-89Arg mutant is shown in green and the charge swap mutant Lys33Glu/Asp87-89Arg is shown in green. The amide region shows no significant overall difference, suggesting that the global fold is basically identical. A few sharp peaks in the green (Asp87-89Arg mutant) are most likely contaminants. Samples were not isotopically labeled and prepared in 30 mM KP buffer with 30 mM KCl and 1 mM DTT at pH8.0 in 10% D_2O. Spectra were acquired at 25 °C.

A.8. ASD code

Below is the code for the solvent weight calculations:

```
1  #!/usr/bin/python
2  '''
3  Written by Micha BA Kunze (mbakunze@gmail.com) 2014
4
5  Purpose:
6  This programme returns "weights" of solvent closenes: by default
       it will sum up the weights using a 1/r and 1/r^6 weighting.
       These are calculated for a given set of atoms (selection in
       SELECTIONS section). These have to specified using the
       MDAnalysis slections, e.g "segid A and resname ILE and name CD1
       ".
7
8  Usage:
9  ./script.py <pdb-file.pdb> <grid-spacing (A)> <output_name>
10
11 *** The selection of atoms for which to calculate the weights can
       be controlled by the selection string in the SELECTION section
       (at the varaible 'selection').
12
13 Other variables for tuning:
14 n_box_1d : number of domains in 1D of the of the solvent cube
       resulting in n_box_1d^3 domains
15 r_cut : cut off for weight summation
16 all_atoms : use different segid (chain) if you wish
17 ca_atoms : use different segid (chain) if you wish or include
       hydrogens - will take longer
18
19 Dependencies:
20 - sys, numpy, MDAnalysis; mandatory and must be installed
       beforehand
21 - time, datetime; optional
22
```

```
23 MDAnalysis (http://pythonhosted.org/MDAnalysis/index.html)
24 '''
25
26 import sys
27
28 time_b=True
29 try:
30     import time
31 except ImportError:
32     print 'Could not import "time" package - Consider installing.'
33     time_b=False
34
35 try:
36     import numpy as np
37 except ImportError:
38     print 'Could not import NumPy. Plase install first.'
39
40 try:
41     import MDAnalysis
42 except ImportError:
43     print 'Could not import MDAnalysis. Please install first.'
44     sys.exit()
45
46 datetime_b=True
47 try:
48     import datetime
49 except ImportError:
50     print 'Could not import datetime package. \nConsider
          installing, programme will continue without.'
51     datetime_b=False
52
53 #start benchmark timer
54 if time_b is True:
55     start=time.time()
56 #VARS
```

```
57 #first sys variables
58 pdb=sys.argv[1]
59 grid_spacing=float(sys.argv[2])
60 out_fname=sys.argv[3]
61
62 #more control variables
63 #min_rad should be changed to mimick the PRE agent size (center(
       PREagent)-center(atom))
64 #r_cut is the sum radius cut-off and serves as a volume
       normalizatin therefore
65 #r_cut can be increased but will result in more grid points ->
       linger computation
66 #n_box_1d can be tweaked depending on protein size
67 # -> resulting box_l should be > 2*min_rad
68 # -> len(domain_ndx[key]) should be > ~20 for efficiency (
       guesstimate)
69 # read SETUP QUALITY CHECK output
70 min_rad=float(3.5)      # minimal distance for solvent criterium
71 r_cut=float(30.0)       # distance until where to sum up
72                         # r_cut < (box_size-protein size)
                               !!!!!!!!!!!! ATT for big guys
73                         # box size now set on r_cut
74 n_box_1d=14             # INT no of boxes in one dimension n_box_1d
       ^3 total number for DD
75 xyz_out=1              # INT controls type of output 0-4, default:1
        (ONLY imprint domains.xyz printed)
76
77 #test vars FOR DEBUGGING ONLY
78 a=0 #a=b=c=0 safe for verbose >=1 mode
79 b=0 #but prob. not inside exclusion radius
80 c=0
81 verbose=0 # INT controls output level 0-3
82 #empirical adjusments /mem allocation - try not to touch
83 corr_box=1.0001 # needed to get n_box_1d to be the real box total
       (FUDGE)
```

```
84 dt=np.float16 # half precision floats (can be float64 - but takes
       loads of RAM) and is usually unnecessary
85 #START
86 print '\n#############################'
87 print 'Called %s %s %s %s'%(sys.argv[0], sys.argv[1],sys.argv[2],
       sys.argv[3])
88 if datetime_b is True:
89     st = datetime.datetime.now()
90     print 'Time: %s'%st
91
92 #SELECTIONS
93 print '\n#############################'
94 print 'Loading structure %s'%pdb
95 print 'Atom selections can be customized in the SELECTIONS section
       .'
96 print 'r_cut and solvent_radius (i.e. min_rad) can also be set.'
97 u = MDAnalysis.Universe(pdb)                                    #
       load structure
98 c_alpha=u.selectAtoms("name CA and segid A")                    #
       for COM calc.
99 all_atoms=u.selectAtoms("protein and segid A and not name H")   #
       for making imprint
100
101 #select you atom positions here
102 selection=u.selectAtoms("resname ILE and name CD1 and segid A") #
       ACTUAL SELECTION OF ATOMS
103
104 #CALCULATIONS AND SETUP
105 #calc. COM of protein
106 print '\n#############################'
107 print 'Start of the calculation.'
108 print 'Calculating COM of structure'
109 total=len(c_alpha.coordinates()[:,1])
110 com=np.zeros(3)
111 for coord in c_alpha.coordinates():
```

```
112     com=com+np.divide(coord,total)
113
114 print 'Calculating exclusion radius'
115 r_tmp=[]
116 for coord in all_atoms.coordinates():
117     r_tmp.append(np.linalg.norm(coord-com))
118 r_exclude=max(r_tmp)+min_rad     # where to calc the imprint
119 box_size=2*(r_exclude+r_cut)     # total box length for iso
120 fine=int(box_size/grid_spacing) # number of lateral/axial points (
        iso)
121                                  # now set depending on spacing and
                                         r_cut
122 grid_points=np.power(fine,3)     # total number of solvent gri
        points
123 points=np.linspace(-0.5*box_size,0.5*box_size,num=fine) #
        construction vector
124
125 box_l=np.divide(float(box_size),float(n_box_1d)) # domain side
        length (iso!)
126 box_edge2=np.sqrt(3)*box_l # com-com distance for 3D next neighbor
127
128 if verbose>=1:
129     print 'Domain side length: '+str(box_l)
130     print 'D-D neighbor distance: '+str(box_edge2)
131
132 domain_ndx={}        #list of ndx for points in domain
133 domain_neighbors={} #neighbor list of domain
134 domain_com={}        #domain com
135
136 #QUALITY CHECK PRINTS
137 #check domain length is at least double of min_rad to ensure
        correctnes of neighbor list calculations
138 print '\n############################'
139 print 'SETUP QUALITY CHECKS:'
140 if box_l <= 2*min_rad:
```

```
141     print '***ATTENTION*** \nBAD box_length/solvent_radius ratio:
            '+str(np.divide(box_l,min_rad))
142     if box_l <= 1.01*min_rad:
143         print 'This will lead to empty boxes, i.e. addtional
                overhead from the domain decomposition.\nProgram will
                terminate...'
144         sys.exit() # empty domains are inefficient and cant lead
                to KeyErrors
145
146     print 'Consider reducing no. boxes in 1D (n_box_1d) below: ' +
            str(np.ceil(np.divide(box_size,2*min_rad)))
147     print 'OR decreasing the grid spacing below: '
148 else:
149     print 'Domain_length/Solvent_radius ratio: '+str(np.divide(
            box_l,min_rad))
150     print 'OK'
151
152 #check how coarse the model is
153 grid_rat=np.divide(float(grid_spacing),float(min_rad))
154 if grid_rat >= 1.0:
155     print '***ATTENTION*** \nBAD grid_spacing/solvent_radius -
            could lead to strongly discretized results: '+str(grid_rat)
156     print 'Consider choosing a grid_spacing below: '+str(0.5*
            min_rad)
157 else:
158     print 'Grid_spacing/Solvent_radius ratio: '+str(grid_rat)
159     print 'OK'
160
161
162 print '\n###################################'
163 print 'GRID BUSINESS:'
164 print 'Building a grid with spacing of %0.1f (A)' % grid_spacing
165 print 'Box dimensions: %0.1f %0.1f %0.1f (A)\n'%(box_size,box_size
    ,box_size)
166 next=-1 #counter for building the grid with loops
```

```
167 grid=np.zeros((grid_points,3),dtype=dt)
168 print 'pre-allocated grid (MEM USAGE)'
169 for x in range(fine):
170     for y in range(fine):
171         for z in range(fine):
172             next=next+1
173             grid[next,:]=[points[x],points[y],points[z]]+com
174             #print grid[next,:]
175             x_slice=int((points[x]-points[0])/(corr_box*box_l))
176             y_slice=int((points[y]-points[0])/(corr_box*box_l))
177             z_slice=int((points[z]-points[0])/(corr_box*box_l))
178             try:
179                 domain_ndx[x_slice,y_slice,z_slice].append(next)
180             except KeyError:
181                 domain_ndx[x_slice,y_slice,z_slice]=[]
182                 domain_ndx[x_slice,y_slice,z_slice].append(next)
183                 domain_com[x_slice,y_slice,z_slice]=np.asarray([
                        points[0]+com[0]+((float(x_slice))*box_l)+0.5*
                        box_l, points[0]+com[1]+((float(y_slice))*box_l
                        )+0.5*box_l,points[0]+com[2]+((float(z_slice))*
                        box_l)+0.5*box_l])
184
185
186 if verbose>=2:
187     print '\nISO SLICE CHECK: '+str(0.5*box_l)
188     for key in domain_com:
189         print '\n box id: '+str(key)
190         print 'box com: '+str(domain_com[key])
191         print 'box ndx len: '+str(len(domain_ndx[key]))
192
193 print 'Boxing up the struture coordinates'
194 #box-up protein
195 domain_pr_coord={}
196 for coord in all_atoms.coordinates():
```

```
197     x_slice=int(np.divide((float(coord[0])-float(points[0]+com[0])
            ),(float(corr_box*box_l))))
198     y_slice=int(np.divide((float(coord[1])-float(points[0]+com[1])
            ),(float(corr_box*box_l))))
199     z_slice=int(np.divide((float(coord[2])-float(points[0]+com[2])
            ),(float(corr_box*box_l))))
200
201     try:
202         domain_pr_coord[x_slice,y_slice,z_slice].append([float(
                coord[0]),float(coord[1]),float(coord[2])])
203     except KeyError:
204         domain_pr_coord[x_slice,y_slice,z_slice]=[]
205         domain_pr_coord[x_slice,y_slice,z_slice].append([float(
                coord[0]),float(coord[1]),float(coord[2])])
206
207 #box-up selection
208 domain_sel_coord={}
209 for ci, coord in enumerate(selection.coordinates()):
210     x_slice=int(np.divide((float(coord[0])-float(points[0]+com[0])
            ),(float(corr_box*box_l))))
211     y_slice=int(np.divide((float(coord[1])-float(points[0]+com[1])
            ),(float(corr_box*box_l))))
212     z_slice=int(np.divide((float(coord[2])-float(points[0]+com[2])
            ),(float(corr_box*box_l))))
213
214     try:
215         domain_sel_coord[x_slice,y_slice,z_slice].append([float(
                coord[0]),float(coord[1]),float(coord[2]),selection[ci
                ].segid,selection[ci].resid,selection[ci].name])
216     except KeyError:
217         domain_sel_coord[x_slice,y_slice,z_slice]=[]
218         domain_sel_coord[x_slice,y_slice,z_slice].append([float(
                coord[0]),float(coord[1]),float(coord[2]),selection[ci
                ].segid,selection[ci].resid,selection[ci].name])
219
```

```
220 #print domain_sel_coord
221 print 'Constructed a %i x %i x %i grid' %(fine,fine,fine)
222 print 'Using a %i x %i x %i domain decomposition' % (n_box_1d,
    n_box_1d,n_box_1d)
223
224 print '\nCreating protein imprint in solvent grid:'
225 #make neighbor dict
226 for key in domain_ndx:
227     for key2 in domain_com:
228         if key != key2:
229             com_com=np.linalg.norm(domain_com[key]-domain_com[key2
                ])
230             if com_com <= corr_box*box_edge2: #corr_box is fudge
                factor for numerics
231                 try:
232                     domain_neighbors[key].append(key2)
233                 except KeyError:
234                     domain_neighbors[key]=[]
235                     domain_neighbors[key].append(key2)
236
237 #make r_cut neighbor dict
238 domain_rcut_neighbors={}
239 for key in domain_ndx:
240     for key2 in domain_com:
241         if key != key2:
242             com_com=np.linalg.norm(domain_com[key]-domain_com[key2
                ])
243             if com_com <= corr_box*(0.5*box_edge2+r_cut): #
                corr_box is fudge factor for numerics
244                 try:
245                     domain_rcut_neighbors[key].append(key2)
246                 except KeyError:
247                     domain_rcut_neighbors[key]=[]
248                     domain_rcut_neighbors[key].append(key2)
249
```

```
250 # check for which domains are close to protein for imprinting
251 # i.e. check if domain is within (iso) protein radius
252 # if not skip and do nothing for the points in the domains
253
254 imprint_domains=[]       #list of all domains close to the protein
255 if verbose>=2:
256     print domain_ndx[a,b,c]
257
258 for key in domain_ndx:  #key identifies the domain that is checked
259     com_dist=np.linalg.norm(domain_com[key]-com)
260     if com_dist >= (r_exclude+0.5*box_edge2):# can be excluded?
261         if verbose>=1:
262             print 'nothing to do for domain %i %i %i' %(key[0],key
                    [1],key[2]) # yeah! less work
263     else:
264         imprint_domains.append(key) #domain needs to checked
                against atom coords
265
266 def domain_search(key,p):#calculates distance from grid point to
    protein in the SAME domain
267     dist=float(2*min_rad)
268     if key in domain_pr_coord:                    #protein in this
            domain?
269         for coord in domain_pr_coord[key]:
270             dist=float(np.linalg.norm(grid[p,:]-coord))
271             if dist <= min_rad:
272                 domain_ndx[key].remove(p)         #remove grid
                        point (i.e. too close to protein)
273                 return p, 1                       #return index
                        and flag that p has been removed
274     return p, 0                                   #return index
            and flag that p needs to be checked
275                                                   #against
                                                          neighboring
                                                          domains
```

```
276
277 def neighbor_search(key,p):#calc. distance from grid point to
        protein in neighboring domains
278     for neighbor in domain_neighbors[key]:
279         if neighbor in domain_pr_coord:
280             for coord in domain_pr_coord[neighbor]:
281                 dist=float(np.linalg.norm(grid[p,:]-coord))
282                 if dist <= min_rad:
283                     domain_ndx[key].remove(p)
284                     return                          #if found and
                            removed return early
285     return                                          #not found ...
            still alive
286
287 waiter=0 #progess reporter
288 for key in imprint_domains:
289     waiter=waiter+1
                                                        #progress
            reporter
290     print '\nChecking domain: '+str(key)
                                      # ''
291     print '--> '+str(waiter)+' out of '+str(len(imprint_domains))
            # ''
292     run_list=domain_ndx[key][:] # NEED TO SPLICE COPY THE LIST
            !!!!!!!!!
293                                 # not splicing leads to skipping
                                        of points - deepcopy would also
                                         work
294     for pp in run_list:
295         po, gp = domain_search(key,pp) #same domain search
296         if gp == 0: #if not found in same domain continue with
                neighboring ones
297             neighbor_search(key,po) # neighbor domain search
298
299 #ANALYSIS SECTION
```

```
300 ##########################
301 def domain_r(key,coord): # calculates the distance to solvent
        points in the same domain
302     r_list=[]             # returns a list of all distance values
            below r_cut
303     for i in domain_ndx[key]:
304         r=np.linalg.norm(coord[:3]-grid[i,:])
305         if r <= r_cut:
306             r_list.append(r)
307     return r_list
308
309 def neighbor_r(key,coord): # calculates hte distance to solvent
        points in r_cut-neighbor domains
310     r_list=[]               # returns a list of all distance values
            below r_cut
311     for nkey in domain_rcut_neighbors[key]:
312         for i in domain_ndx[nkey]:
313             r=np.linalg.norm(coord[:3]-grid[i,:])
314             if r <= r_cut:
315                 r_list.append(r)
316     return r_list
317
318 waiter=0
319 r1={}
320 r6={}
321 for key in domain_sel_coord:
322     waiter=waiter+1
323     print '\nAnalysing weights in domain: '+str(key)
324     print '--> '+str(waiter)+' out of '+str(len(domain_sel_coord))
325     for coord in domain_sel_coord[key]:
326         r_same=domain_r(key,coord)
327         r_neighbor=neighbor_r(key,coord)
328         r_list=[r_same+r_neighbor]
329         r1[coord[3],coord[4],coord[5]]=np.sum(np.power(r_list,-1))
330         r6[coord[3],coord[4],coord[5]]=np.sum(np.power(r_list,-6))
```

```
331
332 #OUTPUT SECTION
333 out_file=open(out_fname,'w')
334
335 #make header
336 out_file.write('# solvent weights (1/r) (1/r^6) \n')
337 out_file.write('# parameters: \n# %0.1f grid-spacing (A) \n# %0.1f
        r_cut \n'%(grid_spacing,r_cut))
338 out_file.write('# chain_id | resid | atom | 1/r | 1/r^6\n')
339 for key in r1:
340     out_file.write('%s %i %s %f %f \n' %(key[0],key[1],key[2],r1[
          key],r6[key]))
341 out_file.close()
342
343 # EXTRA AND VERBOSE OUTPUT SECTION
344 ...
```

A.9. Video 1

Video 1 is available @ https://vimeo.com/69111840

Video 1 shows the L1 and L2 loop region during the molecular dynamics simulation of the free HDAC8. The states as characterised in the main text are indicated during the video. The simulation structure is shown in cyan, which is overlaid with the 'binding state structure of the crystal structure PDB 2V5W shown (green) and the crystal structure of PDB code 1T64 (red). These overlays show how the L1 loop structures differ.

A.10. Video 2

Video 2 is available @ https://vimeo.com/106485813

Video 2 shows the L1 and L2 loop region during the molecular dynamics simulation of the the HDAC8:SAHA complex. SAHA is rendered in a cyan stick representation. The simulation structure (cyan) is overlaid with the crystal structures of PDB code 2V5W (green) and PDB code 1T64 (red).

A.11. Video 3

Video 3 is available @ https://vimeo.com/106485816 [un-published, password: mbak2014].

Video 3 shows a slice of HDAC8 (grey surface representation) during an 'in' transition of a water molecule through the transient tunnel. Boxes shown correspond to the boxes used in the analysing algorithm described in the main text. Total simulation time elapsed in the Video is 800 ps.